

## Experiments in publishing

**Nature's web forum on access to the primary literature has highlighted the risks, as well as the attractions, in enforcing availability without charge to readers.**

“Our idea — a rabbit out of the hat — of institutional repositories, will make the university library system sit up and listen.” The rabbit described by Ian Gibson, who is a UK member of parliament, is a recommendation that funding bodies require scientists to make their articles more accessible by publishing free copies on their institutions' websites. It came from the House of Commons Science and Technology Committee in July, after a five-month inquiry into journal pricing, chaired by Gibson.

The US National Institutes of Health (NIH) now seeks to do likewise. It plans to ask researchers to copy final versions of their manuscripts onto the US National Library of Medicine's freely accessible PubMed Central repository six months after publication (see page 115). Those who are concerned about such matters have about two months in which to comment (see <http://grants1.nih.gov/grants/guide/notice-files/NOT-OD-04-064.html>).

But such grand schemes merit serious reality checks. The impact of the NIH plan on the viability of journals and on non-profit learned societies is potentially serious, and questions as to how it might affect the various sorts of journals have been insufficiently explored.

Over the past months, dozens of scientists and publishers have given their views on open access in a web forum at [www.nature.com/nature/focus/accessdebate](http://www.nature.com/nature/focus/accessdebate). One conclusion of the forum, which wraps up this week, is that societies and publishers must remain financially healthy if they are to be able to maintain the quality of information, launch new journals and innovate electronically.

Contributions also indicate that one-size-fits-all solutions are ill-suited to the huge range of journals and business models, from those with low operating costs, low rejection rates and low added value to more selective journals with high costs and significant added value. Is six months too short for some to sustain subscriber revenues?

The forum reveals that many scientific publishers, in particular learned societies, are already experimenting with ways of increasing access. Some give authors the option of paying for their own articles to be made open access. Many already make articles available from their own websites without charge after a delay. Given the latter, might it not be more sensible to encourage access to free material on publishers' sites, for example, instead of creating a parallel universe of papers on a single US-government-sponsored repository?

Some open-access advocates see open-access archiving as a step that would eventually oblige all journals to adopt a single business model of 'author-pays'. In this model, instead of covering the costs of publishing primary papers from subscriptions paid by readers, costs are met by charging fees to authors, their institutions or sponsors.

Visionaries predict that free institutional archives will cause libraries to cancel subscriptions, cutting journals' revenues and forcing them to turn to either author-pays or other open-access models, or downsize to become no-frills providers of peer-review services, or both. The author-pays model is being explored by open-access publishers such as the Public Library of Science and BioMed Central, but neither has yet demonstrated that it can break even, let alone that the model is sustainable for the entire literature.

*Nature*, founded on the traditional publishing model, welcomes alternative models that deliver enhanced access, provided that they also foster added editorial and publishing value. Its publishers are experimenting with new publishing models. The *Nature* forum indicates that the viability of new models remains far from established.

For politicians or state institutions in effect to support one business model would be risky. Diverse approaches and competition are usually preferable, particularly if investment and deployment of technological and other innovations are to be encouraged. ■

## Helping depressed children

**The FDA should improve public information and work to create a publicly accessible clinical-trials database.**

The US Food and Drug Administration (FDA) will meet next week to consider how to calm the firestorm over the use of antidepressant medications in children (see page 122). It is now armed with much more material than it had at its last meeting on selective serotonin reuptake inhibitors (SSRIs) in February. None of these data have altered the conclusions drawn by FDA epidemiologist Andrew Mosholder ahead of the February meeting, however. As a class, SSRIs are associated with an increased risk of suicidal thoughts and behaviours in young people.

The FDA is reluctant to discourage the use of SSRIs, because there are few good treatments for childhood depression, which can by itself lead to suicide. But there is little to suggest that children are better off on the drugs than not. Only one of the drugs, Prozac, has consistently improved childhood depression more than placebo treatment.

The FDA needs to make a statement reflecting these facts, to counteract the long holiday that drug companies have enjoyed. Crucial data about the lack of efficacy and side effects of SSRIs in children have

not been clearly described or even published in medical journals. Without access to the full findings, US doctors wrote 10.8 million prescriptions for the drugs for children and adolescents in 2002. The FDA is responsible for stopping this widespread 'off-label' use.

Drug companies are now allowed to market their products directly to US consumers. But they do not reveal all the information they have about these drugs to potential patients. A new method of disclosure must be created that is as understandable as the TV and magazine ads that spread rosy messages about today's blockbuster drugs.

The drug companies have begun to respond to such demands by posting information about approved drugs on their websites. But the industry's reputation is so tarnished that this is unlikely to be sufficient. The FDA and other agencies should collaborate to develop a clinical-trials database containing information on every clinical trial sponsored by a drug company. It should include all the clinical results from drugs that have been approved for any indication. Such a move would go some way towards restoring the FDA's dented credibility. ■

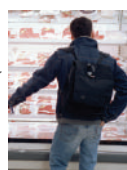




**Culture clash**  
Lack of cash  
threatens to hurt  
Europe's heritage  
p114



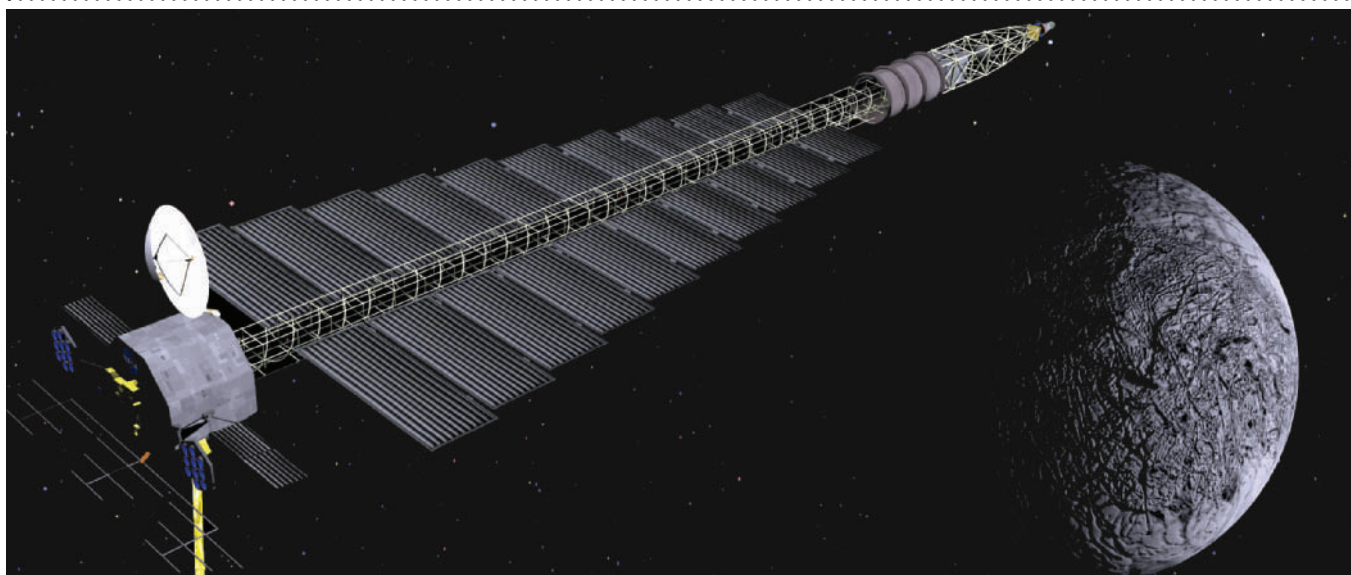
**Free for all**  
US biomedical agency  
prepares for open  
publication  
p115



**Bad taste**  
Consumer review  
falls foul of food  
chemist's ire  
p117



**Logged on**  
Exhibition counts  
the contribution of  
slide rules  
p119



Up in the air: NASA is struggling to fund the development of the nuclear-powered craft proposed for the JIMO mission to Jupiter's moons.

# Reviewers caution NASA over plans for nuclear-powered craft

**Tony Reichhardt, Washington**

An ambitious NASA project to use nuclear reactors to power spacecraft may be heading for trouble, based on early reactions in Congress and from scientists studying future space missions that would depend on them.

Later this month, NASA will choose an industrial contractor to begin designing its first such spacecraft, the proposed Jupiter Icy Moons Orbiter (JIMO). The mission, tentatively planned for launch around 2015, has strong personal backing from the agency's administrator Sean O'Keefe.

NASA surprised the scientific community last year by proposing JIMO, which would orbit several of Jupiter's moons at close range, and would be more technically capable than any scientific spacecraft ever built.

But many scientists are concerned about the mission's cost and political viability. NASA won't even hazard a guess on the final price tag until definition studies are completed next summer, says Ray Taylor, who heads Project Prometheus, the broader nuclear programme that includes JIMO. But estimates to build the nuclear propulsion system alone run as high as \$4.5 billion, and the spacecraft structure and science instruments would add billions more.

NASA officials claim that some of the upfront investment in designing JIMO will make future nuclear missions cheaper. But either way, the project promises to consume a significant fraction of the agency's current annual space science budget of \$4 billion.

JIMO would unquestionably be a big step up in technology. Its nuclear reactor would provide at least 10,000 watts of electricity for scientific instruments, against the 300 watts available to the Cassini spacecraft now orbiting Saturn. The power boost would allow sophisticated equipment such as ice-penetrating radars to fly for the first time and would greatly increase the volume of data from cameras and other instruments.

But a National Academy of Sciences panel, asked by NASA to look at the value of nuclear power for space missions, met in Washington on 31 August to 2 September and seemed underwhelmed by presentations from NASA managers on possible follow-on missions to JIMO. A Neptune mission, for example, would still take nearly 20 years to return scientific data, offering no real advantage over conventional missions. And it would require much bigger rockets than the largest planned US launch vehicle to deliver the craft to Earth orbit before it headed into deep space.

Weighing the costs and long travel times against the scientific return, planetary scientist Reta Beebe of New Mexico State University in Las Cruces said at the meeting that the nuclear missions could be a hard sell with scientists. "This is not inspiring," she said.

Jerry Grey, a space-policy expert with the American Institute of Aeronautics and Astronautics who recently chaired a space nuclear-power round table in Washington, says that building JIMO poses no technical showstoppers, but "it ain't gonna be easy". It will require a thorough ground testing programme, and the risk of failure will still be high. Virtually everything on the mission will be new technology that will need to run for a decade or more without breaking down in space.

Daunting as these challenges are, JIMO is also off to a bad start politically. This summer the House appropriations committee cut \$230 million from next year's budget request of \$438 million for Project Prometheus. Unless the money is restored by the Senate, the effect on the programme would be "severe", says Taylor, who claims that Congress supports the project, and had to cut its funding because of competing budget demands. "To our knowledge, there's a tremendous amount of support" for it going ahead, he says. ■



## Peer-reviewed paper defends theory of intelligent design

Jim Giles

A new front has opened up in the battle between scientists and advocates of intelligent design, a theory that rejects evolution and is regarded by its critics as another term for creationism.

A scientific journal has published a paper that argues in favour of intelligent design — the first time such material has appeared in a peer-reviewed publication, according to biologists who track the issue. The paper appeared in a low-impact journal, *Proceedings of the Biological Society of Washington*. But critics say that it could still be used by advocates of intelligent design to get the subject on to US school curricula (see *Nature* 416, 250; 2002).

The article comes from the Discovery Institute in Seattle, Washington, a leading promoter of the theory. In the article, senior fellow Stephen Meyer uses information theory and other techniques to argue that the complexity of living organisms cannot be explained by darwinian evolution (S. C. Meyer *Proc. Biol. Soc. Wash.* 117, 213–239; 2004).

Many of Meyer's arguments have already been aired by advocates of intelligent design, but critics say that publication will be used to back up claims that the theory is scientifically valid.

Kenneth Miller, a cell biologist at Brown University in Providence, Rhode Island, who has argued against Meyer in public debates, does not doubt that this will happen. "They've tried very hard to get material into peer-reviewed journals."

Richard Sternberg, a taxonomist at the National Center for Biotechnology Information in Bethesda, Maryland, was editor of the journal publishing the Meyer paper when it was reviewed and accepted. Sternberg is also on the editorial board of the Baraminology Study Group, which publishes papers on "scientific research in creation biology". He says the paper was seen and approved by three well-qualified referees.

Meyer's article has attracted a lengthy rebuttal on *The Panda's Thumb*, a website devoted to evolutionary theory. But Miller says that, despite criticism of the journal, versions of the theory will find their way into the scientific literature at some point. Arguments for it can be written, he says, as reappraisals of certain aspects of evolution rather than outright rejection. "Peer review isn't a guarantee of accuracy," he adds. "That is especially true of review articles." ■

## Conservers plead for funds to protect Europe's heritage

Alison Abbott, London

Tens of millions of tourists visit Europe each year to enjoy its cultural heritage. And if they were aware of it, they would no doubt cheer the work of the scientists who are preserving that heritage — fighting the harm caused by pollution, by the natural processes of decay and by the tourists themselves.

But at a meeting in London from 1 to 3 September, the European network of these interdisciplinary researchers declared that its support is diminishing.

For more than a decade, the European Commission has funded, at a modest level, programmes designed to analyse damage to cultural heritage and to prevent, monitor and repair such harm. The scientists involved draw on disciplines that include laser optics and nanotechnology; they consider all aspects of decaying heritage, from ancient manuscripts to archaeological sites.

The support they have had from the European Commission has fallen from €40 million (US\$48 million) in the fifth framework programme (1998–2002) to €10 million in the current, sixth programme. "We worry that the trend of undervaluing cultural heritage is worsening," says May Cassar, head of the Centre for Sustainable Heritage at University College London, who organized the meeting.

The commission is currently the only source of funding in Europe that is specifically earmarked for generic research into sustaining cultural heritage, although individual groups sometimes win small grants from national sources. As drafting of the seventh programme, which starts in 2007, begins in earnest, the researchers want political leaders to recognize the value of their work to Europe's vast tourism industry.

One project, called Caramel, has led to the development of monitors of inorganic and organic atmospheric pollutants, which include the carbon that is responsible for blackening buildings. These unobtrusive monitors can be placed at key sites, reports Cristina Sabbioni, a physicist at the National Research Council's Institute of Atmospheric Sciences and Climate in Bologna, who is involved in the project. Its scientists have been able to quantify the atmospheric levels of small particles, which deposit carbon more quickly than large particles, during rush hours in tourist cities such as Florence and Seville. And local authorities in Seville have



Frieze frame: Parthenon sculptures cleaned by lasers were displayed at the Olympic Games in Athens.

decided to restrict traffic flow past its main cathedral, which is currently being cleaned, in direct response to the Caramel data.

Another project, called LaserACT, diagnoses degradation in works of art using lasers. Such lasers are also being used to clean sensitive stone, including the west frieze of the Parthenon, which has been scrubbed by technology developed at the Institute of Electronic Structure and Laser, Crete, in collaboration with ESMA, the group responsible for the conservation of the Acropolis.

Now biologists are joining physicists in the conservation cause, the meeting heard. Microbiologists are analysing the colonies of bacteria that form in caves and catacombs where visitors have disturbed steady light and air conditions and caused damage to the artwork and structures. A project called BioBRUSH is researching calcite-forming microorganisms, which may remove salty crusts on stone and lay down a hard calcite layer to protect it from further damage. "There is a long way to go before this technology could be applied with confidence to important monuments," cautions BioBRUSH scientist Eric May of the University of Portsmouth, UK. "We need reliable sources of funds to develop this type of approach," he says. ■



# Biomedical agency floats open-access plan

Geoff Brumfiel, Washington

The main biomedical research agency in the United States has issued a plan that would make the results of all research that it supports freely available shortly after their initial publication in the scientific literature.

On 3 September, the National Institutes of Health (NIH) posted the plan on its website. It would require all grantees to place their final versions of accepted manuscripts — or the published version itself if the publisher agrees — on the US National Library of Medicine's PubMed Central database within six months of publication.

Comments are invited within 60 days on the plan, which broadly complies with some advisory language that was added to the NIH's budget bill by a congressional committee in July. The NIH is expected to decide on the version to be implemented shortly after that. Details of the plan's implementation will be closely scrutinized by scientific publishers.

"We are aware of many of the implications of possible changes," NIH director Elias Zerhouni told a meeting of representatives of scientific societies and patient advocacy groups at the agency's main campus in Bethesda, Maryland, on 31 August. "At the same time, the mission of the NIH includes delivering research information to the public."

The plan acknowledges, however, that this mission must be "balanced with the ability of journals and publishers to preserve their critical role in the peer review, editing



Open door? The National Library of Medicine's PubMed Central database could keep copies of all NIH-supported research.

and scientific quality control process".

But some publishers oppose the plan, complaining that they have not been sufficiently consulted on its impact on the established scientific communication system. "This is being done rather stealthily," complains Allan Adler, head of government affairs at the Association of American Publishers in Washington DC. "There's been no real discussion of what this is going to cost the taxpayer, or of its impact on publishers."

"Zerhouni is making us all do this experiment," says Martin Frank, executive director of the American Physiological Society in Bethesda, Maryland. "The federal government is trying to regulate the dissemination of information at the expense of an established, diverse publishing operation."

Advocates of open access strongly support the proposed policy change. "I applaud what he's doing," says Harold Varmus, former

director of the NIH and president of the Memorial Sloan-Kettering Cancer Center in New York City. "To hide NIH research behind high subscription fees is not fair," he says.

Varmus says that the six-month delay will allow publishers to continue generating income from subscriptions. But Samuel Kaplan, a microbiologist at the University of Texas Medical School in Houston and chair of the publishing board of the American Society for Microbiology, says that won't be the case, particularly for smaller journals that publish quarterly. "It could be the death-knell of society publica-

tions," he predicts.

If implemented in its current form, the NIH plan will be an important victory for advocates of open access to the scientific literature. In July, the House of Commons Science and Technology Committee advised the UK government to encourage universities there to make all their papers freely available online (see *Nature* 430, 390; 2004).

Richard Roberts, a geneticist at New England Biolabs in Beverly, Massachusetts, says that open access is rapidly gathering momentum. On 26 August, Roberts, Varmus and 23 other Nobel laureates co-signed a letter calling on Congress and the NIH to provide open access to government-funded research.

But Frank pledges that the scientific societies will do what they can to amend the NIH's proposal. "We won't roll over," he says. "We will continue to fight this."

♦ <http://grants1.nih.gov/grants>

## Ocean fix for climate change finds tentative support

Jim Giles

Marine organisms can sense and avoid high concentrations of carbon dioxide, according to a study of a seafloor vent off the coast of Hawaii. The result provides tentative support for plans to tackle climate change by dumping carbon dioxide in the ocean.

Researchers have long been concerned that adding high concentrations of CO<sub>2</sub> to the ocean might cause serious damage to local marine life — some studies have shown that it can kill marine organisms such as nematodes (see *Nature* 430, 391; 2004). Environmentalists have blocked some plans to conduct further tests, fearing that even small injections of CO<sub>2</sub> might open the door to larger tests or industrial projects.

So the researchers turned instead to studying a natural plume of CO<sub>2</sub> that

bubbles up from a subsea volcano called Loihi, near Hawaii. They wanted to assess fears that adding CO<sub>2</sub> to the ocean might create a 'mortality sink' — a spot where marine organisms die, attracting scavenging creatures that would in turn be killed.

But this kind of death trap is unlikely to occur, says Jeffrey Summers, a physicist with the Office of Fossil Energy at the US energy department in Washington DC. Summers and colleagues set cages baited with mackerel close to the Loihi plume and at various distances from the CO<sub>2</sub>. The bait away from the plume was eaten in less than 24 hours, whereas the bait over the vent remained untouched for more than a week.

Eric Vetter, a marine biologist at Hawaii Pacific University who worked with Summers on the project, thinks animals

are avoiding the cages because they can sense the high CO<sub>2</sub> levels. "The results are promising," he says.

The study, scheduled to be presented on 6 September at the 7th International Conference on Greenhouse Gas Control Technologies in Vancouver, Canada, also suggests that sea creatures can recover from short blasts of CO<sub>2</sub>. Summers' team dragged cages of amphipods — shrimp-like creatures — over the vent. The animals seemed to be anaesthetized by the gas within 10 minutes, but became active again around half an hour after being removed from the plume.

Vetter stresses that the work is "very preliminary", and adds that much more data are needed before conclusions can be drawn about the wisdom of dumping CO<sub>2</sub> in the sea.

# Suppliers step in to manage chemical use

Jim Giles

The consumption of chemicals in some US industries is being reduced by making suppliers fully responsible for the use and recycling of their products, says a report out next week.

The Chemical Strategies Partnership (CSP), a San Francisco-based organization that promotes this new way of working, has conducted a survey of the approach's spread. It says that the industry has begun a long-term shift away from just selling chemicals towards arrangements in which suppliers are far more actively involved in managing the use of their products. The trend is likely to lead to better environmental stewardship, the CSP says.

The approach is being adopted by large industrial customers in the United States. Rather than just buying paint, for example, car manufacturers such as General Motors are bringing suppliers into their assembly plants to manage the entire painting process.

"The supplier gets paid per car body coated," explains Kevin Swift, an economist at the American Chemistry Council, an industry body in Arlington, Virginia. "So the incentive is for them to cut down on paint."

The CSP's survey, scheduled for publication on 14 September, says that the practice is booming: the market for chemical-manage-



Net paint: chemical suppliers' advice has reduced car factory emissions.

ment services has grown by about half since 2000 and is now worth US\$1.2 billion. Car makers lead the way, with 80% of the industry outsourcing management of chemicals. Aerospace and electronic firms are also using the approach. And the report suggests that, in the long term, academic labs could buy some of their chemicals this way.

General Motors has benefited from the approach, says Michael Knoblock, an environmental services manager at the company's offices in Pontiac, Michigan. In the 1980s, the firm used 10,000 different chemical companies to supply it with the 150,000 or so different substances it needed for its business, he

says. Now it uses about half as many chemicals, sourced from just five firms. "In the first year of implementation we saw 25% savings," says Knoblock. "And all of our management programmes have reduced emissions of chemicals."

Tom Bierma, an environmental health expert at Illinois State University in Normal, says that research laboratories could eventually benefit from the inventory, safety and training services now being offered by suppliers, even though the amount of chemicals they use is relatively small.

This is backed up by CSP case studies of chemical management in US labs. In a study at the Stanford Linear Accelerator Center in

California, it suggested that the lab could save money by coordinating its purchasing, which is currently spread across many individuals in different departments. The CSP also says suppliers could help labs to design more environmentally friendly means of dealing with chemical waste.

The CSP wants to export its approach and there are signs that European companies may be ready to take it up. Knoblock, for example, says that he expects the first chemicals-management programme at a General Motors plant in Europe to be up and running within a year.

♦ [www.chemicalstrategies.org](http://www.chemicalstrategies.org)

## China increases share of global scientific publications

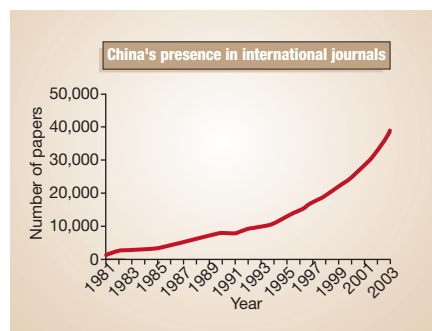
David Cyranoski

China's scientific growth is keeping up with its meteoric economic rise, according to figures released last week. But the growth seems to vary widely between fields and the quality of the work may be lagging in some of them.

Between 1981 and 2003, China clocked a 20-fold increase in its publications in international scientific journals (see graph), reports *Science Watch*, a newsletter tracking trends in scientific activity, published by Philadelphia-based Thomson ISI.

Although China publishes only half as many papers as Japan, it is pulling away from other nations in the region, such as South Korea, Taiwan and Singapore. China now accounts for over 5% of the world's scientific publications.

Papers with authors affiliated to Chinese institutions now account for 10% of the literature in materials science, and for 8% of that in mathematics and physics.



Ron Shen, a materials scientist at the University of California, Berkeley, who is originally from Shanghai, says that China has strong historical foundations in metallurgy. He says its interest in materials science has accelerated with access to better equipment.

Shen notes, however, that papers in many fields, including physics, chemistry and geosciences, still have low 'average impact factors'. The impact factor is a rough measure of how influential a paper is.

"Chinese researchers are good in analytical and technical skills," he says, "but perhaps weaker on reasoning and interpretation."

China remains one of the smaller players in the life sciences, with its share of total publications ranging from 0.8% in immunology to 2% in plant science, *Science Watch* finds. And Mu-Ming Poo, who heads the Institute of Neuroscience in Shanghai, says that some of the overall growth in the nation's scientific profile can be attributed to the thousands of papers from labs in the United States that are co-authored by visiting Chinese researchers and students who maintain affiliations in their homeland.

Chinese authors do account for an increasing number of the world's most influential papers, finds *Science Watch*. But to get ahead, Shen says, China needs to assert itself as a leader and not a follower. "It is still very much influenced by the outside world," he says, "so it follows the research trends of the West very closely."





Where's the beef? *Consumer Reports'* studies of pesticides and irradiated meat have come under fire.

## US chemist attacks consumer magazine's food safety work

**Jim Giles**

A food chemist has accused a leading consumer publication of using flawed science in its food safety surveys.

The researcher says that *Consumer Reports*, which has four million subscribers, ignored published research that contradicted its tests on food irradiation, and used an unscientific toxicity scale to measure pesticide residues.

Joseph Rosen, who works at Rutgers University in New Brunswick, New Jersey, made the claims at a session on organic food at the American Chemical Society's annual meeting in Philadelphia on 23 August. The magazine rejects the allegations, saying its surveys are based on rigorous science developed by professional toxicologists and statisticians.

Rosen says he first looked into *Consumer Reports'* methods after reading an article on irradiated meat, which it published in August 2003. This examined US food producers' claims that treating meat with beams of electrons or  $\gamma$ -rays kills harmful bacteria. It concluded that treated meat had a "slight off-taste" and offered no safety benefit to the "careful cook".

The magazine regularly examines industry claims, and its defenders say that it is a counterweight to the food industry's marketing. The magazine is "highly effective" in its approach, says Caroline Smith-Dewaal, food safety director at the Center for Science in the Public Interest, a Washington-based group that campaigns on food science issues.

But Rosen says that *Consumer Reports* based its irradiation article on reports from just two trained taste testers, and ignored a 2003 test involving more than 100 subjects conducted by researchers at Kansas State University. That study found that

consumers could not tell irradiated from non-irradiated meat.

Consumers Union, the New York-based advocacy group that publishes *Consumer Reports*, says that it knew of the Kansas research, but that the magazine does not typically run consumer panels. "We evaluate attributes and not preferences," says Urvashi Rangan, a toxicology and environmental health expert with the union. She adds that the article stated that the off-taste was subtle and might be missed by untrained tasters.

Rosen also says it is alarmist for the magazine to state that the energy in irradiation is "150 times the dose capable of killing an adult", as consumers are not exposed to the beams. Rangan counters that the measure was needed to correct food industry promotional material, which portrays the beams as no more powerful than sunlight.

The magazine's work on pesticides also aroused Rosen's ire. A March 1999 article used a specially designed toxicity scale to compare residues on fruits and vegetables. *Consumer Reports* concluded that parents should avoid giving children large amounts of some produce.

Rosen says the evaluations used arbitrary factors, and that the scale confuses the maximum safe amount of a residue that can be consumed at once with that which can be consumed over a lifetime. Rangan concedes there is an arbitrary element to the weightings, but says that they were needed to allow comparison of residues in different foods.

"I was struck by the shoddiness of their work," says Rosen, who advises the American Council on Science and Health, a lobby group generally supportive of the food industry. He says he gets no research money from the organization. ■

## Only pride hurt as predicted quake fails to strike California

**David Cyranoski**

Time has run out for a researcher's prediction that a large earthquake would hit California by 5 September. Vladimir Keilis-Borok had predicted that such an event would occur at some point in the past nine months. But, fortunately for residents, nothing of the kind happened.

Keilis-Borok, a geophysicist at the University of California, Los Angeles, hasn't given up. "We can use the mistakes to improve our methodology," he says.

Keilis-Borok has had some success in the past. In 2003, his team made forecasts that were fulfilled by the Tokachi-oki earthquake in Japan on 26 September, and the Californian San Simeon earthquake on 22 December.

Recently he predicted that an earthquake of magnitude 6.4 or greater would hit somewhere in 32,000 square kilometres of southern California (see map). The United States Geological



Survey and the California Earthquake Prediction Evaluation Council said the method "appears to be a legitimate approach". But they took no action; a nine-month window cannot be used to make decisions about evacuation.

Keilis-Borok looks for patterns in the smaller earthquakes that precede large ones. He uses a similar method to analyse other systems such as crime waves and elections; his numbers predict that George W. Bush will win in November.

But some experts think Keilis-Borok's successes have been down to luck. Based on earthquake history, there is a 30% chance of one hitting Tokachi-oki during any nine-month period, although there is only a 2–5% chance of one in San Simeon. Tom Jordan, head of the Southern California Earthquake Center in Los Angeles, says Keilis-Borok's method uses complex algorithms that will have to be more thoroughly evaluated before critics are won over. ■



## Cloning paper pulled from publication after media circus

**London** Fertility researcher Panos Zavos has had a peer-reviewed paper on human cloning pulled — because he publicized his work before publication. The journal's editor says he also has concerns that what Zavos told the press doesn't match the claims in the paper.

Zavos announced at a London press conference last week that he had created cloned embryos by mixing genetic material from dead people with cow eggs. The insertion of human DNA into animal eggs, which are more easily available than human eggs, has been done before. But Zavos' use of dead human subjects has sparked fresh worries that some people might see cloning as a way to bring back the deceased. Zavos has made several controversial cloning claims but has yet to support them with published work.

Zavos submitted a paper on a similar topic to the *Journal of Assisted Reproduction and Genetics*, where it was accepted for publication. But editor-in-chief Norbert Gleicher, who would not disclose details of the paper, says he has concerns about the veracity of Zavos' public portrayal of the



**Panos Zavos claims to have made cloned embryos using DNA from dead humans and cow eggs.**

work. Gleicher says the journal gave the paper's authors two days to comment on these concerns, but received no response.

## Fault drillers try to get to the bottom of Chi-Chi quake

**Tokyo** Seismologists in Taiwan celebrated a promising achievement for earthquake studies this August after drilling into the fault zone responsible for the disastrous 1999 Chi-Chi earthquake.

**The Chi-Chi quake is particularly interesting to scientists as it caused a huge displacement of land — in some places, one**

side of the fault ended up 10 metres higher than the other. The samples taken from the drill hole should shed light on the physics of earthquakes, and help researchers work out how factors such as rock porosity and water content affect the movement of land. "This is a unique opportunity to validate the model proposed from seismic data with direct evidence," says Kuo-Fong Ma, a seismologist at the National Central University in Chungli, Taiwan, who is heading the project.

The researchers have so far struck four different fault lines at different depths down to 1,240 metres. They will continue to drill to 2,000 metres, as well as drilling a second bore hole, over the next year. Other researchers are drilling into other faults in search of similar data, including the San Andreas fault in California.

## Splatometer is a big hit in driving bug survey

**London** Car drivers are collaborating with wildlife researchers in an attempt to monitor changes in the number of insects in Britain — by counting the bugs that are splattered on their licence plates.

The Splatometer survey was thought up by researchers at the Royal Society for the Protection of Birds (RSPB). "People had called us to say that they were finding fewer

## Slide rules line up to mark times past

**Washington** There was a time when no self-respecting scientist would be seen without the ultimate in computing technology: a slide rule.

An exhibition at Purdue University in West Lafayette, Indiana, is taking a nostalgic look at the history of these computing tools, which ruled from their invention in 1632 until the 1970s.

"There was a point in time when the slide rule was king," says James Alleman, a civil engineer who has spent 15 years collecting slide rules from Purdue alumni.

Alleman, shown here (on the right) with retired civil engineering professor Robert Miles, proudly displays one of their prize specimens, which



measures more than two metres in length. The collection includes a rule once owned by astronaut Neil Armstrong.

"If these slide rules could talk ... I am sure they would tell remarkable stories," says Alleman.

dead insects on their windshields during the Sunday car wash," says Caroline Osborne of the RSPB. "We needed to quantify that drop because we were worried about the birds that feed on them, especially the house sparrow communities, which have fallen by 65% in 31 years," she says. Researchers speculate that pesticides and habitat loss may be causing insect numbers to decline.

During June, 40,000 volunteers counted 324,814 bugs, which hit plates at an average rate of one splat every 8 km. The society plans to repeat the survey in 2006 to check for changes.

## US centres set up to ponder broader issues of genetics

**Washington** The ethical, legal and social questions surrounding genetics and genomics are set to be tackled by four new centres in the United States, funded by \$20 million from the US National Human Genome Research Institute (NHGRI).

Elizabeth Thomson, director of the NHGRI's programme on these issues, says the plan marks a new direction for the institute, which has typically funded smaller, investigator-led research in these matters.

The centres will be established at Stanford University, California; Case Western Reserve University in Cleveland, Ohio; Duke University in Durham, North Carolina; and the University of Washington in Seattle. Together, they will gather expertise from many disciplines, such as law, bioethics, science and theology.

## Singapore opens door to stem-cell research

**Tokyo** Singapore has become the latest country to ban the use of cloning for reproductive purposes, but to allow the work for the development of stem cells.

The law, passed on 2 September, prohibits several practices that might be used in human cloning, including placing cloned human embryos into the wombs of humans or animals, and allowing cloned embryos to develop for more than two weeks. But the restrictions do not prevent the development and harvesting of stem cells from cloned human embryos.

Only one company currently does research with embryonic stem cells in Singapore, according to health minister Balaji Sadasivan. But the country is working hard to build up its biotechnology industry, and observers speculate that the business will grow in the wake of this legislation.

# The bolt catchers

Every summer, Florida plays host to researchers who fire rockets at the sky to create lightning. And having captured the bolts, the group has found something shocking, as Mark Schrope finds out.

The early Greeks saw it as a weapon of the gods, and throughout history lightning has been viewed almost universally as something best avoided. But at a desolate patch of sand and grass in northern Florida, a handful of researchers gather each summer to call down lightning from the sky in the name of science.

Lightning strikes are unpredictable, and the research is turning out that way too. Amid studies to find fresh ways to protect homes and businesses from the wrath of storms, scientists examining the basic properties of lightning were struck by a bolt from the blue. For decades, it would seem, physicists have been working with incomplete models of these electrical discharges.

Key to the success of the studies is a high natural occurrence of thunderstorms — and a rocket launcher. In the 1960s, researchers showed that lightning can be generated in a controlled fashion by firing a rocket, trailing a grounded wire, at a thundercloud.

“Even though lightning is something you don’t want to mess around with, the advantage of triggered-lightning studies is that you can actually capture lightning to measure its currents,” says Paul Krehbiel, a lightning researcher at the New Mexico Institute of Mining and Technology — known as New Mexico Tech — in Socorro.

Scientists at flash points around the globe now use the rocket-and-wire technique to create and study lightning. Florida boasts some of the highest strike rates for natural lightning on the planet, and at the International Center for Lightning Research and Testing at Camp Blanding military base, scientists have triggered lightning during each thunderstorm season for the past 11 summers.

At the heart of the 40-hectare facility, operated by the University of Florida in Gainesville, is a 12-metre wooden tower topped by eight rocket launchers. Scattered around the tower are detectors that measure when the surrounding electric field is high enough for lightning strikes. As the facility’s co-director Martin Uman explains, the electric field he and his team look for before launching a rocket just exceeds that which

halts rocket and shuttle launches by NASA.

When conditions are right, a technician in a nearby trailer packed with monitors triggers the launch of a one-metre rocket trailing 700 metres of Kevlar-coated copper wire grounded to cables on the tower. The system succeeds about half the time with the wire literally burning up in a blaze of lightning glory.

Much of the physics involved is still poorly understood, but in basic terms lightning begins when excess negative charge builds up in a cloud and creates a downwardly moving spark invisible to human eyes. Called a step leader, this spark can reach the ground in about 20 thousandths of a second. The large negative charge in the spark induces positive charge on the

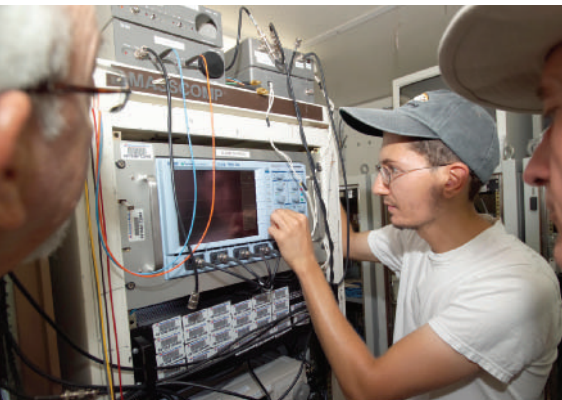


No flash in the pan: researchers are generating their own lightning (above) to gain insight into bolts from the blue.

ground as it approaches, creating sparks that move upwards. Once a step leader meets an upward spark, a continuous path between cloud and ground is formed along which charge from the cloud is dumped to create the luminous display we know as lightning.

The conductive wire carried by the rockets essentially reverse this process — as do skyscrapers at times — by initiating an upwardly moving spark that reaches a charged zone above. Once a channel is formed, the wire disintegrates under the blistering temperatures (between 8,000 °C and 33,000 °C) of the discharge, and the lightning itself is essentially the same as a natural bolt.

This summer, Uman’s team is studying systems for protecting buildings and power lines from lightning. The group directs the



electricity from triggered lightning through cables to a lightning rod attached to a newly built house just a few metres from the tower and to a full-size power-line system.

## Protect and survive

The rod is part of a commercial lightning-protection system, which aims to divert most of the lightning’s current away from the house to the ground. This protection system has generally performed as designed, but in all the triggered experiments the currents recorded in the house have exceeded theoretical expectations. “The measurements so far look very good,” says Uman, “but we don’t understand them completely, which means that it’s not the simple situation everyone would like to think.”

Uman explains that most protection systems are based on laboratory experiments





and theory, rather than on studies of lightning itself. But customers need to know how well the equipment works in practice. For example, power companies want to protect their power lines from the disruption caused by lightning but they need to weigh the benefits against the cost of upgrading their protection systems.

When lightning hits the house, detectors at the site measure each bolt's electric field, and digital cameras capture the strikes from every angle, providing data for fundamental research. Uman and the researchers at Camp Blanding work closely with scientists from other institutions, such as Joseph Dwyer of the Florida Institute of Technology in Melbourne.

In 2002 Dwyer began tackling a question that lightning researchers have struggled with since the 1920s: does lightning produce X-rays? The debate has rumbled on for decades, although researchers at New Mexico Tech published tantalizing results in 2001 that strongly indicated that X-rays are emitted near natural lightning<sup>1</sup>. Dwyer believes that triggered lightning offers the potential to answer the question once and for all.

Dwyer and his colleagues measure X-rays using off-the-shelf radiation detectors held inside heavy, welded aluminium cases that block out electromagnetic interference from thunderstorms. The equipment also features fibre-optic connections to protect against stray incoming electrical signals.

The team first deployed the detectors near the rocket launch tower in 2002 and, after

studying 26 triggered bolts, became convinced that lightning reliably produces X-rays<sup>2</sup>. Dwyer says that the data were so decisive and repeatable that it is now clear that X-ray emission is a major characteristic of lightning omitted from previous models. "There is something fundamentally wrong with our understanding of how lightning works if we can miss such a big ingredient," he says.

In general, the emissions seem to be associated with the upward sparks and possibly the very beginning of the bright lightning stroke. They come in brief microsecond-long bursts, with energies comparable to those used to take dental X-rays.

### Shock to the system

But in summer 2003, the group's last rocket-triggered bolt of the season revealed something even more incredible than the X-ray emissions.

An unusually strong current pulse arrived at the detector before the lightning was fully under way. Such spikes happen occasionally, but are not well understood.

During this early spike, the instruments recorded hundreds of  $\gamma$ -rays 10 to 300 times more powerful than the X-rays they had been observing<sup>3</sup>. The team believes that these rays originated several kilometres above the ground, based on the time that they arrived at the detectors. Such intense  $\gamma$ -rays could possibly be detected from space if anyone were looking. They might even explain

radiation bursts recorded by Earth-orbiting satellites during thunderstorms.

Because of atmospheric absorption, the intensity of  $\gamma$ -rays at the detector is far lower than would be found in the cloud. But the exact point at which the bursts originate has yet to be confirmed. This year, Dwyer and his team hope to pinpoint the source by using modified detectors that are shielded so that only rays from a single direction can get in.

It is not yet clear what causes either the X-rays or the  $\gamma$ -rays, but Dwyer leans towards an unproven atmospheric process called runaway breakdown, first suggested in 1961. Subatomic particles such as electrons are peculiar in that the faster they move, the more the drag acting on them is reduced. The idea is that electrons could be accelerated by strong electric fields inside a thundercloud or near lightning, ultimately approaching the speed of light because of reduced drag. As these high-energy electrons collide with air molecules they could create yet more electrons, eventually generating the X-rays and  $\gamma$ -rays.

### Storming ahead

Dwyer also believes that runaway breakdown could play a role in causing lightning itself. Conventional wisdom says that lightning is initiated when charge builds up in clouds to such an extent that the associated electric field overcomes air's insulating ability. The resulting breakdown allows current to flow through the air, creating the first spark. But this would theoretically require an electric field larger than has ever been measured in association with lightning. Runaway breakdown, in contrast, could involve an electric field ten times smaller, which is closer to the measured values.

We can expect lightning research to accelerate if work planned for later this year at New Mexico Tech proves successful. An international collaboration plans to test the idea that laser pulses can be used to trigger

lightning by ionizing air to create discharge paths. Historically, all tests of this theory have failed, but the group will be using a terawatt laser — the most powerful laser ever applied to the problem. Lasers are more expensive than rocket launchers, but laser triggering

would allow almost unlimited attempts at lightning generation during thunderstorms.

Even without such power, Dwyer feels lucky to be doing lightning research. "Usually there are so many people in a field that you have to spend years doing something to make your mark," he says, "but we can make a discovery with just about every measurement."

**Mark Schroppe is a science writer based in Florida.**

1. Moore, C. B., Eack, K. B., Aulich, G. D. & Rison, W. *Geophys. Res. Lett.* **28**, 2141–2144 (2001).
2. Dwyer, J. R. et al. *Geophys. Res. Lett.* **31**, L05118 (2004).
3. Dwyer, J. R. et al. *Geophys. Res. Lett.* **31**, L05119 (2004).



# Bitter pills

They were hailed as wonder drugs to banish depression, but may cause suicidal thoughts in some children. Companies are under attack for failing to publish these data. Erika Check considers the science behind a scandal.

In a packed hotel ballroom on a cold February day, 30 parents stepped one by one to a microphone to testify to the US government. They had come to suburban Maryland from as far away as Texas and Arizona to talk about antidepressant drugs. Some had brought their children. But others could bring only photos.

Mark Miller said that his son, Matt, ended his life after taking seven doses of the antidepressant Zoloft, made by Pfizer. Glenn McIntosh told how his daughter, Caitlin, hanged herself after taking Zoloft and another drug, GlaxoSmithKline's Paxil. A few parents spoke up in favour of the drugs, arguing that they had helped their children. But others echoed McIntosh's anguished plea: "We were told that antidepressants like Paxil and Zoloft were wonder drugs, that they were safe and effective for children. The pharmaceutical companies have known for years that these drugs could cause suicide in some patients. Why didn't we?"

It's a question that has shaken Americans' faith in the companies that provide their medications, and in the regulatory system that is supposed to ensure that drugs work, and

are safe. Although most of the drugs have not been licensed for use in children, doctors have widely prescribed them 'off label' to withdrawn and depressed kids. So why, ask campaigners, were regulators and the public not told the results of clinical trials that cast doubts on the wisdom of prescribing the drugs to children?

In this emotionally charged atmosphere, the US Food and Drug Administration (FDA) will meet on 13 and 14 September to consider the scientific evidence about giving drugs called selective serotonin reuptake inhibitors, or SSRIs, to children. After

reviewing the available data in February, and hearing the parents' testimony, the agency told manufacturers to label SSRIs with warnings that ask doctors to watch patients closely for suicidal tendencies. It also commissioned a new evaluation of the clinical data, drawing on the assistance of researchers at Columbia University in New York.

## Depressing evidence

The results of this analysis are now in, along with findings from other studies. It seems clear that SSRIs are, in some cases, linked to suicidal behaviours in children. But what this link means — and what regulators should do about it — remains extremely contentious. Does the risk apply to all drugs in the class? Does it exceed the risk of suicide in children whose depression is left untreated? How effective are any of the SSRIs in relieving depression in children? And, more fundamentally, has the controversy exposed worrying gaps in our understanding of childhood psychiatric conditions?

There are no simple answers, but Britain's Medicines and Healthcare Products Regulatory Agency has



Pill power: children are prescribed a range of antidepressants.

J. RAEDLE/GETTY IMAGES

J. RAEDLE/GETTY IMAGES





Up in arms: this year's American Psychiatric Association meeting saw protests at child medication.

already seen fit to advise that SSRIs apart from Prozac, made by Eli Lilly, should not be given to children under 18. Regulators in many other countries are looking to the FDA for leadership. The media are scenting a scandal, and activists are on the warpath (see 'Big pharma's *bête noire*', overleaf). With children's lives and mental well-being at stake, the FDA faces its sternest test in years.

After Prozac was released onto the market in the late 1980s, there was a scare that it caused suicidal impulses in some patients. But in 1991, the FDA ruled that there was no clear evidence, and gave the drug a clean bill of health. The controversy about SSRIs blew up again last year, when the FDA asked GlaxoSmithKline to provide information on its unpublished trials of Paxil in depressed children. When those data revealed disturbingly high incidences of suicidal thoughts and behaviours, the agency demanded the results from 24 drug-company studies on SSRIs in children.

Many of the studies, conducted by eight different firms, had never before been published or presented to the public. When FDA epidemiologist Andrew Mosholder combined the data from the trials, he found that children who received SSRIs were 1.89 times as likely than those on a placebo to show suicidal thoughts and behaviours.

But there were doubts about how well these behaviours had been classified. So the FDA asked psychiatrists at Columbia University to reclassify the data. On 20 August,

the agency released the results of the reanalysis, which confirmed Mosholder's central finding: children on SSRIs were 1.78 times as likely to exhibit suicidal tendencies.

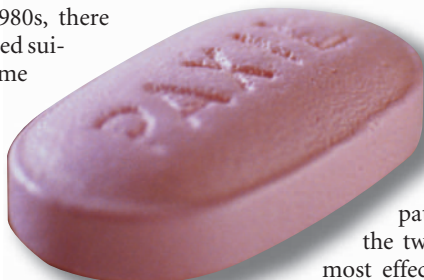
More evidence on the drugs' side effects arrived last month from the Treatment for Adolescents with Depression Study (TADS), led by John March of Duke University Medical School in Durham, North Carolina. His team studied two treatments for childhood depression: Prozac, and cognitive-behavioural therapy, which encourages depressed people to change their thought patterns. A combination of the two seemed to provide the most effective relief of depression.

But children given Prozac, in isolation or together with behavioural therapy, were more likely to report harming or thinking about harming themselves or others<sup>1</sup>.

### Cure and kill?

Critics of SSRIs have leapt on the latest findings. "I think it's pretty clear that the drugs cause a problem," says David Healy of the University of Wales in Bangor, who has persistently attacked drug companies for their promotion of SSRIs.

But why should drugs designed to alleviate depression tip some children over into suicidal behaviour? Some psychiatrists blame a side effect called 'akathisia' — an agitated state that occurs in a minority of patients placed on antidepressants<sup>2</sup>. Others argue that children starting drug therapy might become vulnerable as their depression



lifts enough for them to voice or act on impulses that they felt before the treatment began. "In primary care, we've been taught for years that when depressed patients start to feel better, they can get enough energy to commit suicide," says David Price, a psychiatrist based in Denver, Colorado, who writes guidelines on the treatment of depression for the healthcare provider Kaiser Permanente.

Evidence from a large study of depressed adults and children published in July seems to support this theory, showing a peak in suicidal tendencies within nine days of starting treatment with antidepressants<sup>3</sup>. "We don't really know, but the medicine probably makes you less likely to stop yourself," says March.

Nevertheless, none of the children in any of these trials actually killed themselves, and the incidence of behaviours related to suicide was low — in the TADS trial, for instance, only about 10% of the children given Prozac reported harming or thinking of harming themselves or others. So many psychiatrists still argue that the drugs are safer than no treatment at all.

### Trials on trial

"Depression is a serious illness that needs treatment," says Graham Emslie, a child psychiatrist at the University of Texas Southwestern Medical Center in Dallas, who has consulted for manufacturers of SSRIs. "And if you want to use the most effective treatment you've got to use the medicine."

The problem with this argument, however, is that Prozac is the only SSRI to have been shown to alleviate childhood depression in rigorous clinical trials, and to be approved by the FDA for that purpose. Nevertheless, doctors routinely prescribe other SSRIs to depressed children, with Zoloft and Paxil topping the list in the United States. Other popular SSRIs include Wyeth's Effexor and Celexa, made by Forest Pharmaceuticals. Because the drugs are considered safe and effective for depressed adults, doctors have simply assumed that the same must be true in children.

Indeed, if Prozac helps depressed children, it's difficult to understand why other SSRIs should not. SSRIs are designed to work by increasing the availability of a neurotransmitter called serotonin, which affects mood and emotions. Seeing no reason to question the underlying pharmacology, many experts have assumed that the clinical trials in depressed children are flawed. Some psychiatrists point out that most SSRIs show some trend towards working in children, even if this is not statistically significant. "It's not the case at all that a failed trial disproves efficacy," says Robert Findling of Case Western Reserve University in Cleveland, Ohio, who has also served as a paid consultant for SSRI manufacturers.

One of the biggest problems with testing antidepressant drugs in children is that



young people are especially likely to show a strong placebo response, which may mask the drugs' biological effects. In one of the successful trials of Prozac, for instance, the children given dummy pills improved 70% as much as those on the drug itself<sup>4</sup>. The attention that depressed children gain simply from being part of a clinical trial may help to explain these results. "Even if a kid gets a placebo, he still has to show up at a clinic, and someone will ask him questions about how he's feeling," says Christopher Varley, a psychiatrist at the University of Washington at Seattle.

Another possible problem is that trials of depressed children may include some individuals with bipolar disorder, formerly known as manic depression, for which SSRIs are not an effective treatment. People with bipolar disorder often don't experience their first manic phase until early adulthood, which could mean that the patients enrolled in adult and paediatric trials of SSRIs may actually have very different clinical profiles. "We could be comparing apples with oranges," says Price.

### Proper treatment

While supporters of SSRIs use these arguments to justify their belief that the drugs help children, for others they underline the need for caution. If we don't know enough about childhood depression to accurately diagnose it, and if the attention focused on children in a clinical trial can be enough by itself to improve their condition, then might it be best to avoid drugs?

Jon Jureidini, a child psychiatrist at the University of Adelaide in Australia, believes the real problem is that mental-health services often fail to provide behavioural therapies for children who need them, which leads to an over-reliance on drugs. "The danger is that we're masking this problem by drugging children, which may be quieting them down, but may also be indirectly harming them," he says.

Jureidini also takes little comfort from

## Big pharma's *bête noire*

On Sunday 18 July, Congressman James Greenwood's office gave Vera Hassner Sharav some bad news: a hearing on antidepressants in children had been cancelled.

Sharav, an activist for patients' rights who has become a thorn in the drug industry's side, was livid. She was even more upset when she learned that Greenwood had cancelled the hearing because he was stepping down from Congress to take a new job at the Biotechnology Industry Organization, of which many antidepressant manufacturers are members. So Sharav e-mailed her distribution list of journalists, politicians and other opinion formers, accusing Greenwood of "selling out" to the drug industry.

Greenwood, a moderate Republican from Pennsylvania, was so stung that he called Sharav to demand that she distribute his rejection of her accusation, and his assurance that the hearing would be rescheduled — which she did, in a brief statement, on 21 July. But by then, an article quoting Sharav's criticism of Greenwood had already appeared in *The Washington Post*.

The incident demonstrates Sharav's methods, and her growing influence. For those following the controversy surrounding the prescription of antidepressants to children, Sharav's partisan 'infomails' have become a constant narrative.

Her exhaustively compiled mailing list includes cabinet members, government officials and scores of journalists.

Sharav began investigating the ethics of clinical research and the drug industry more than a decade ago, after her son died from a reaction to a drug used to treat schizophrenia. After his death, Sharav found out that the prescribing doctor was on retainer for a company conducting research on the drug.

Sharav, a law librarian, began using her research skills to look for cases of what she considered to be ethical abuses in clinical research. Eventually, this became a full-time occupation, and in 2001 she formed the non-profit Alliance for Human Research Protection.

Initially, Sharav was perceived as an outsider. But she has gradually built up her influence by sending her infomails, and by organizing for families to testify at key government hearings. "As events began to catch up with what I was saying, I was proven to be right, and that's what has lent increasing support to the cause," she claims.

Pharmaceutical executives shouldn't expect Sharav to rest on her laurels. "I don't consider this battle in any way over," she says.

♦ [www.ahrp.org](http://www.ahrp.org)

the relatively low rate of suicidal behaviours reported in the drug companies' clinical trials. Such trials are short, he says, and are not designed to pick up rare events such as suicide attempts. So the fact that a consistent risk of suicidal tendencies appears across all of the paediatric SSRI trials is worrying, Jureidini argues. "We know from our experience with other drugs that harms often become apparent after the trial period is well and truly over," he says.

Most of the psychiatrists who doubt the drugs' efficacy don't want to see them

banned for use in children. The main problem, they say, is that they are being prescribed too often by family doctors who have little knowledge of mental health, who don't support drug treatment with behavioural therapy, and who typically don't monitor their patients closely to look for the signs of suicidal impulses. Prescribing SSRIs to adults may now be routine, says Price, but for children it shouldn't be. "I'd advise a primary care doctor to call a psychiatrist when diagnosing depression in a child or adolescent," he says.

The drugs' manufacturers are now changing their labels according to the FDA's instructions, and reject the suggestion that they concealed data. Eli Lilly and GlaxoSmithKline have also announced plans to place data from trials on approved drugs on the web. A spokeswoman for GlaxoSmithKline told *Nature* that the firm is eager to see how the FDA interprets the childhood SSRI trial results. "It is not a black and white issue," she says.

The stakes are high for the FDA. It must make sense of the evidence and take appropriate action — or it might have to answer more questions from devastated parents mourning their lost children. ■

**Erika Check is *Nature's* Washington biomedical correspondent.**

1. March, J. et al. *J. Am. Med. Assoc.* **292**, 807–820 (2004).
2. Hamilton, M. S. & Opler, L. A. *J. Clin. Psychiatr.* **53**, 401–406 (1992).
3. Jick, H., Kaye, J. A. & Jick, S. S. *J. Am. Med. Assoc.* **292**, 338–343 (2004).
4. Emslie, G. J. et al. *J. Am. Acad. Child Adolesc. Psychiatr.* **41**, 1205–1215 (2002).

E. VUCCI/AP



Public grief: parents of allegedly drug-damaged children at a hearing on the issue in February.

# Science priorities ignore Colombia's water needs

The government did not consult with the scientific community on reforms.

*Sir*—Colciencias, the Colombian agency responsible for funding scientific research, is trying to overcome critically low investment in science and technology by establishing scientific research consortia in six strategic areas. These 'centres of research excellence' will have a budget of US\$1.7 million per centre over the next five years. The current call for proposals, which ends on 24 September, sadly ignores research related to water resources and the environment. This is a major omission, but our efforts to correct it have been in vain.

For the most part, the selected areas are relevant and genuine priorities, including biodiversity and genetic resources; infectious tropical diseases; energy development; biotechnology and food innovation. However, critical areas — most notably water and the environment — are being ignored in a

country facing urgent challenges. Such omissions are made more striking by the inclusion of nanotechnology in the list of national priorities: a field unlikely to flourish here, given the lack of expertise.

Issues such as water availability and supply, water-resource management and planning, are continuing challenges in Colombia. The provision of safe drinking water is an immediate necessity for our children, dozens of whom suffer and die every day as a result of drinking polluted water. Intense tropical storms trigger flash floods and avalanches along Colombian rivers; droughts claim an even higher toll.

According to IDEAM (Colombia's public agency for hydrology, meteorology and environment), the deterioration of the Colombian environment is accelerating. The interaction between natural and social

systems in the tropical Andes is one of our most urgent research challenges.

Perhaps most worrisome is the process by which the Colciencias list was drawn up. Hundreds of leading scientists and scientific institutions in Colombia, including ours, did not participate in any way in its creation. The priorities discussed in government internal documents were not circulated publicly, let alone reviewed thoroughly by the wider scientific community. It is essential that Colciencias actively involves the community in these processes, to guarantee their legitimacy and relevance for the country, in both scientific and social terms.

**Germán Poveda**

*Graduate Programme in Water Resources, School of Geosciences and Environment, Universidad Nacional de Colombia, Medellín, Colombia*

## Complexity of the body calls for animal research

*Sir*—The criticism of the UK animal regulatory system by Dan Lyons in Correspondence ("The animal-care regulatory system is a sham", *Nature* **430**, 399; 2004) calls for a response. Currently in my fiftieth year of pharmaceutical research, I find it depressing to realize how many legislators and citizens accept the unsupportable position of the animal activists regarding the need for animals in biomedical research.

In my lectures and writings on this subject, I emphasize the unbelievable complexity of the intact animal body. Each day, the human heart pumps some 7,200 litres of blood through the trillions of cells in the body on a continuous basis. The cells and organs share biochemical messengers every second of the day. Despite all the information published on the function of the animal body, my guess is that we know only a small percentage of the reactions going on at any moment. What conceivable system could be devised outside the intact animal that would be able to mimic the complexity of the animal body?

I agree that animals should be properly housed and maintained and suffering reduced as much as it is humanly possible to do so. Also, every surrogate test system (cell culture, enzymes, genomics, proteomics and so on), along with computers, should be and are used daily in the laboratories.

But what logical human would agree to have a new chemical, never before tested

in a whole animal, administered for the first time to himself or herself?

If the animal activists prevail, new drug discovery and development will cease. This at a time when we have available to us the greatest number of disease targets in history. Why can we not get this message across to our citizens and legislators? Where is the activists' protocol for discovering a new drug without using animals? Should they not be required to detail a programme in which animals will not be used?

**Charles G. Smith**

*USA (full address supplied)*

## Classroom volunteers inspire young ecologists

*Sir*—Your News Feature "Doing it for the kids" (*Nature* **430**, 286–287; 2004) highlights INSPIRE, a science-education scheme in the United Kingdom where postdocs pursue a teaching qualification while continuing with their research. Although there are many positive attributes to INSPIRE, your feature acknowledges the scheme's heavy reliance on special funding, which may disappear in the future.

As noted in the News Feature, the United States currently lacks programmes equivalent to INSPIRE. But there are other US programmes that achieve many of the same goals with the help of volunteers, and without the need for additional funding. One example of this is Kids Do Ecology (KDE), an outreach programme run through the National Center for Ecological Analysis and Synthesis

(NCEAS) in Santa Barbara, California.

This programme pairs postdocs at NCEAS with classes of 10-year-olds at local schools. Each volunteer scientist visits his or her class about six times during the spring, guiding students through the scientific process by helping them develop a research question and hypothesis, design experimental methods and conduct the experiment. At the end of the semester, students from each class present their projects to other students and the NCEAS community at a poster session. A description of the experiments can be found on each class's web page, hosted through the KDE website ([www.nceas.ucsb.edu/nceas-web/kids](http://www.nceas.ucsb.edu/nceas-web/kids)).

The benefits of KDE are many. It provides scientists with the opportunity to participate in community outreach, gives students some hands-on science experience and introduces a scientist role-model to the classroom. The programme requires a minimal time-commitment from the postdocs, about 20 hours in total on a volunteer basis, and consequently has a small budget. KDE is a sustainable model for getting scientific expertise into the classroom, and participation has remained high since its inception in 1997: 11 scientists visited classes at four Santa Barbara schools this year.

The framework for KDE is simple and effective, and could easily be adopted by other research institutions wanting to inspire student interest in science.

**Sarah Abramson**

*National Center for Ecological Analysis and Synthesis, 735 State Street, Suite 300, Santa Barbara, California 93101, USA*



# Symbol minded

Relationships with others have helped to shape our ability to learn language.

## The First Idea: How Symbols, Language, and Intelligence Evolved From Our Primate Ancestors to Modern Humans

by Stanley I. Greenspan & Stuart G. Shanker

Da Capo: 2004. 512 pp. \$25, £18.99

Peter Hobson

What are the bases for human symbolic thinking and language? This is a question that prompts fierce confrontation between those who defend what is distinctively interpersonal about human social engagement, and those who aspire to computational reductionism in modelling cognitive development. *The First Idea* joins the attack on the side of the interpersonal camp (rather like a Sherman tank, in fact), to shake if not destroy the forces of nativism and individualism.

The impetus behind the intellectual assault comes from a conviction that the origins of symbolizing, both in human prehistory and in the early development of young children, lie in emotionally patterned interpersonal relations. It will not do to pivot an account of the origins of thought and language around cognitive constraints that are hard-wired into the brain. In *The First Idea*, Stanley Greenspan and Stuart Shanker insist that we have radically underestimated the developmental role played by emotions that link humans together. They consider that “successively more complex interactive emotional signals”, and the cultural practices that are founded on and exploit such communication, both structure and integrate the psychological abilities required for thinking.

It takes many pages for Greenspan and Shanker to lay out their theory as they range over the domains of child development, primate social and communicative abilities, autism, the origins of culture, and even the future evolution of humanity. Modesty is not an obstacle to their forthrightness: again and again we are told how, largely through their own observations, the authors have enjoyed flashes of insight denied to others. Yet for all their expansiveness, the authors are justified in thinking that they challenge some widely held presumptions about the nature and development of thinking. Notwithstanding the contributions of Lev Vygotsky and his followers, who have stressed the social origins of higher forms of cognition, it is still radical to situate communication between people not at the periphery of human thinking and language, but at its core. The clinical experience of Greenspan



Play school: interactions between children help them to improve their symbolic thinking.

(a child psychiatrist and psychoanalyst) and the thoughtfulness of Shanker (a philosopher and psychologist) allow them to elaborate a distinctive view of human development and psychopathology.

It is this viewpoint that sets this book apart. Although I found myself resisting the sweep of the authors' 16 developmental stages of emotional and intellectual growth, the book's cumulative effect is to give the reader a deeper appreciation of the power and formative potential of human emotional interaction. It contains enough provocative ideas about the kinds of social and developmental processes at work in various aspects of intellectual and cultural life to set the framework for new theoretical and empirical investigations.

Regarding human evolution, for example, it is highly plausible that it was the increased ability to communicate and coordinate subjective states with others, together with concurrent changes in cultural practices, that led to a revolution in cognitive functioning through the acquisition of symbolic thinking. In the field of developmental psychopathology, the authors have good reasons to view the primary deficit in autism as the child's difficulty in engaging with others on an emotional level. There is something theoretically, as well as therapeutically, compelling in observations that these children's cognitive functioning may improve substantially when it proves possible to enhance their relationships with other people. Just as *The First Idea* expounds how patterns of mutual emotional and communicative exchange with others are vital

for typical development, so too its authors point out that where there is no mutuality, the repercussions are profound.

The devil is in the detail. Often the authors make statements with little more support than their personal observations. For example, they state that the human capacity for engaging in longer and more continuous chains of emotional signals is what allows us “to negotiate and solve problems, and, thereby, more fully separate perceptions or images from their fixed actions and construct higher and higher levels of internal symbols”. They tell us how “emotional recognition that one's actions can have an impact on someone else is the foundation for sequencing, that is, the ability to carry out many steps in a row where each one is related to the previous one”. These are interesting possibilities, but is there evidence to convince us one way or the other? The authors often cite studies in child development, anthropology or neuroscience that seem to be congruent with their position. But do not expect to find detailed or critical analyses of scientific evidence. In flourishing a broad brush, the authors seem intent to impress the general reader rather than engage the mind of the scientific sceptic.

Apart from matters of evidence, does this book really explain how interpersonal engagement yields symbolic thinking? Here the authors chide Piaget for failing to come up with the goods but, other than correcting the one-sidedness of his theory, do they really do better? Take their account of how symbolizing is achieved by separating perception from action. Piaget identified

the critical property of symbolizing as the emancipation of thought from action, and of meanings from the objects to which those meanings usually apply. What seems to be missing from *The First Idea* is an account of the mechanisms by which social influences allow children to ascribe alternative perspectives, recognized as such, to and through symbols. There are candidate explanations — my own view is that the emotionally grounded process of identification between a young child and others does the trick — but the authors seem to feel that their arguments suffice.

Through their creative thinking about emotional and interpersonal aspects of early human development, Greenspan and Shanker have helped us to find our bearings for the intellectual fight ahead. I just wish their map had been adjusted in scale, to something nearer pocket size.

Peter Hobson is at the Developmental Psychopathology Research Unit, Tavistock Clinic and University College London, 120 Belsize Lane, London NW3 5BA, UK. He is the author of *The Cradle of Thought* (Macmillan, 2002).

## Sizing up a growing field

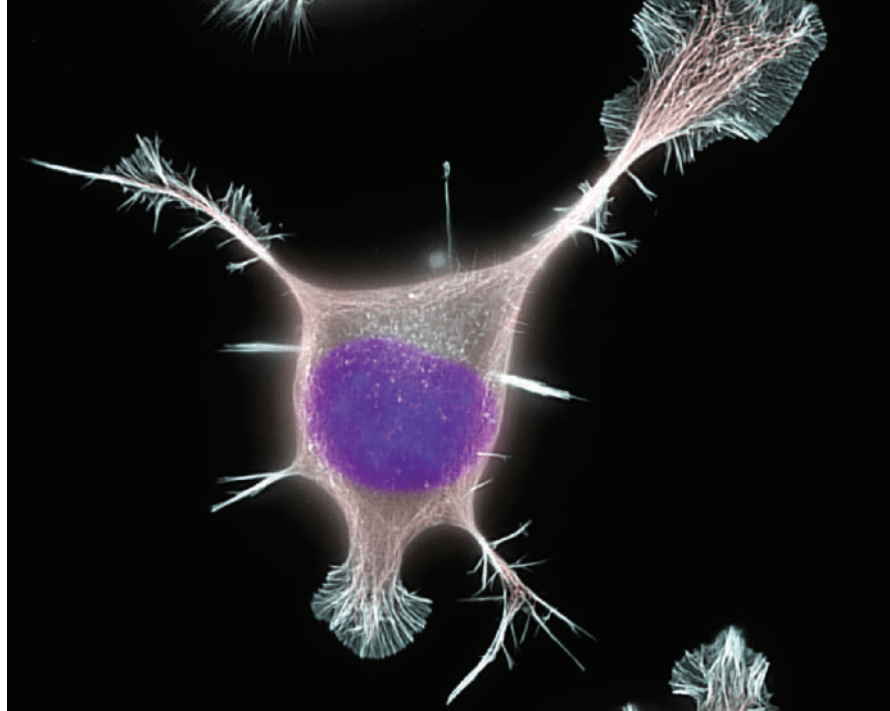
### Cell Growth: Control of Cell Size

edited by Michael N. Hall, Martin Raff & George Thomas  
Cold Spring Harbor Laboratory Press: 2004.  
652 pp. \$135, £95

**Brendan D. Manning and  
Lewis C. Cantley**

The terms 'cell growth' and 'cell proliferation' are often used synonymously to mean an increase in cell number. Strictly, though, cell growth refers to an increase in cell size or mass, whereas cell proliferation is an increase in cell number due to cell division. The two processes together determine the ultimate size of cells, tissues, organs, even organisms. Cell proliferation is almost always accompanied by cell growth, except, for example, in the early stages of embryogenesis. But cell growth often occurs without cell division, for instance in terminally differentiated cells such as myocytes and neurons.

Cell growth underpins many critical cellular and developmental processes, yet studies on its regulation and mechanisms have lagged behind those on cell proliferation and cell-cycle progression. However, a flurry of genetic, cell biological and biochemical studies over the past ten years have made great strides towards discovering the signalling pathways and mechanisms that drive cell growth. *Cell Growth*, edited by three investigators at the forefront of this research, Michael Hall, Martin Raff and



**Growth factor:** neurons can increase their size dramatically by extending axons and dendrites.

George Thomas, details our current knowledge of this field.

The all-star cast of authors assembled for this work gives a clue to the book's likely impact. In the foreword, Paul Nurse sets the stage by defining perhaps the most perplexing question in the field: how does a proliferating cell double its mass and contents once per cell division, with near-perfect precision? Central to this question is whether or not a cell has the capacity to monitor its own size, and several chapters revisit this problem. Nurse's discussion of the growth patterns used to achieve a constant cell size is complemented by Patrick O'Farrell's opening chapter on the evolution of body size in metazoans. O'Farrell brilliantly delineates the constraints that all higher organisms face in

reaching their final size, and the enormous variety of developmental patterns used to overcome these constraints.

Pioneering studies on cell growth control by Nurse and Lee Hartwell in yeast, and by Anders Zetterberg and others in mammalian cells, have focused on the nature of the relationship between cell size and cell-cycle progression. Many proteins and pathways involved in cross-talk between these two processes have since been characterized in yeast, flies and mammals, and these are detailed in the early chapters. The main theme that emerges is the wide range of distinct mechanisms to coordinate growth and division that exist between different systems, even in different developmental stages of the same organism. The wiring of networks

## New and revised textbooks

### Human Evolutionary Genetics: Origins, Peoples and Disease

by Mark A. Joblin, Matthew Hurles & Chris Tyler-Smith  
Garland Science, 523 pp. \$61.95, £35

### Understanding DNA: The Molecule and How it Works (3rd edn)

by Chris Calladine, Horace Drew, Ben Luisi & Andrew Travers  
Elsevier, 334 pp. £29.95, \$45

### Animal Physiology

by Richard W. Hill, Gordon A. Wyse & Margaret Anderson  
Sinauer, 776 pp. \$104.95, £37.99

### Gastrulation: From Cells to Embryos

edited by Claudio D. Stern  
Cold Spring Harbor Laboratory Press, 731 pp.  
\$150, £100

### Introduction to Protein Science: Architecture, Function, and Genomics

by Arthur M. Lesk  
Oxford University Press, 310 pp. £26.99, \$51.95

### Molecular Markers, Natural History and Evolution (2nd edn)

by John C. Avise  
Sinauer, 655 pp. \$59.95, £38.99

### Immunobiology: The Immune System in Health and Disease (6th edn)

by Charles Janeway, Paul Travers, Mark Walport & Mark Shlomchik  
Garland Science, 848 pp. \$74.95;  
Churchill Livingstone, £39.99

### Human Evolution: An Illustrated Introduction (5th edn)

by Roger Lewin  
Blackwell, 277 pp. \$44.95, £19.99



## Science in culture



## Class action

Werner Tübke's huge wall-painting will grace a new building at the University of Leipzig.

**Alison Abbott**

Part of the extraordinary legacy of Werner Tübke, one of the former East Germany's most renowned artists, who died in May, is this vast wall painting in the rectory of the University of Leipzig.

The story of the painting is at least as interesting as the multifaceted story it tells. The university, founded in 1409 by Pope Alexander V, is among the oldest in Europe. Its beautiful central buildings were extensively damaged by bombing raids in the Second World War, however, and the new Communist-style administration building was an ugly concrete block. To inaugurate the building, the government opened up a competition in 1970 for an artwork to reflect the theme: "The working class and intellectuals are inseparably bound in socialism under the leadership of the Marxist-Leninist Party." Tübke won.

His *Arbeiterklasse und Intelligenz* ("Working Class and Intellectuals"), completed in 1973, is a wall painting of immense, floor-to-ceiling proportions, some 13 metres wide. It was



painted with oils and tempura on wood, and it depicts a series of discourses laden with social commentary. Some of the figures are recognizable portraits of contemporary local dignitaries. The immense detail and Renaissance references are typical of Tübke, whose most well known work is a monumental panorama depicting the struggle between peasants and the bourgeoisie in the sixteenth century and includes more than 3,000 figures.

Tübke, who has been referred to as "the last court artist", was never apologetic about accepting contracts from the East German state—and his works have never been simple political endorsements. *Arbeiterklasse und Intelligenz* is so valued that when the university opened a competition last year to design a building to replace the concrete block, one of the criteria was that any design should be able to display this grand work appropriately. The painting's new home is due to be completed in time for the 600th anniversary of the university in 2009.

controlling cell growth is dependent on the cell context, but the pathways that make up these networks are generally shared. Nine of the chapters discuss pathways, such as the 'target of rapamycin' pathway, and processes, such as protein synthesis, that are at the heart of all cell growth. Finally, the book closes with several chapters discussing the specific growth-control mechanisms in specialized cells, such as lymphocytes, myocytes and neurons, which vary in size from one another by many orders of magnitude.

In compiling this book, the editors have successfully overcome many obstacles. For instance, it is clear that cell growth control mechanisms often vary significantly between commonly studied systems. Rather than discount this fact, the editors have put together a complementary collection of chapters that address this problem and search for

common themes. Furthermore, with each chapter representing a separate review, one might expect a high degree of redundancy. Although there is some overlap, the authors of each chapter do a remarkable job of staying focused on their area of expertise yet still manage to tie their subject into the wider context.

Our knowledge of the pathways that control cell growth is increasing rapidly, making it difficult to publish an up-to-date book on the intricacies of this field. However, the overall balance achieved between detailing our current understanding of the molecular control of cell growth and outlining the fundamental research challenges that remain, make this book indispensable to those with new, or renewed, interest in this topic.

Collectively, the different chapters cover a

wide range of topics that will be of particular interest to cell, developmental and evolutionary biologists alike. Some of the mechanistic details described, however, although necessary, may be difficult for a general audience to follow. But the book's multidisciplinary nature will make it an excellent reference for many scientists. *Cell Growth*, like many titles from the Cold Spring Harbor monograph series, should stand the test of time and serve as a solid foundation for this ever-expanding field.

Brendan D. Manning is in the Department of Genetics and Complex Diseases, Harvard School of Public Health, Boston, Massachusetts 02115, USA. Lewis C. Cantley is in the Division of Signal Transduction, Beth Israel Deaconess Medical Center, and the Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, USA.

# Natural proportions

Redfield ratios: the uniformity of elemental ratios in the oceans and the life they contain underpins our understanding of marine biogeochemistry.

Paul G. Falkowski and  
Cabell S. Davis

An interesting empirical observation in biology is the relationship between the elemental composition of organisms and ecosystems. All organisms are composed primarily of a mixture of six major elements: hydrogen, carbon, nitrogen, oxygen, phosphorus and sulphur. But the proportion of these basic ingredients varies between organisms — and such variations can lead to interesting properties within ecosystems.

For example, in the oceans most of the biomass comprises small drifting organisms (plankton) that are rich in nitrogen. These organisms are essentially functionally similar ensembles of metabolites, often encased in a shell formed from the most readily available ingredients. Much plankton is consumed by other plankton with similar chemical compositions. The result is that on average, the nitrogen:phosphorus (N:P) ratios of plankton in the oceans are remarkably similar throughout the world, averaging approximately 16:1 by atoms. When these organisms or their body parts sink into the ocean interior, their energy-rich bodies are consumed by bacteria which, in aerobic conditions, oxidize the organic matter to form dissolved inorganic nutrients, especially  $\text{CO}_2$ ,  $\text{NO}_3^-$  and  $\text{PO}_4^{3-}$ .

In 1934, Alfred Redfield wrote a now classic paper in which he proposed that the N:P ratio of plankton (16:1) causes the ocean to have a remarkably similar ratio of dissolved  $\text{NO}_3^-$  and  $\text{PO}_4^{3-}$ . This hypothesis suggested that, devoid of life, the chemical composition of the oceans would be markedly different. The concept of Redfield ratios has been fundamental to understanding of the biogeochemistry of the oceans ever since.

The basic problem with Redfield ratios is that they are empirical. The ratios were originally derived from measurements of the elemental composition of plankton, and the  $\text{NO}_3^-$  and  $\text{PO}_4^{3-}$  content of seawater from a few stations in the Atlantic, but were subsequently supported by hundreds of independent measurements. Yet there is no known reason why the average N:P ratio of plankton should be 16:1. Why not 6:1? Or 60:1? If one looks at the elemental composition of individual species of phytoplankton grown under nitrogen or phosphorus limitation, the N:P ratio can vary from around 6:1 to 60:1. Redfield understood this problem, but did not try explain it, except to note that the N:P ratio of inorganic nutrients in the ocean interior was an average, and that small-scale



Just drifting along: a phytoplankton bloom.

variability around the mean was to be expected. Despite many reports that the elemental composition of organisms in a region of the ocean does not conform to Redfield ratios, or that the elemental composition of marine phytoplankton grown in cultures is not 16:1, Redfield's fundamental concept remains valid. It cannot be rationalized by reductionist arguments, nor refuted by anecdotal observations. The fact that the  $\text{NO}_3^-:\text{PO}_4^{3-}$  ratio in the interior of all major ocean basins is remarkably similar to the N:P ratio of plankton is due to the residence times of these two elements in the ocean (roughly  $10^4$  years), relative to the ocean's circulation time (roughly  $10^3$  years). As the residence times exceed the mixing times by an order of magnitude, it should not be surprising that the  $\text{NO}_3^-:\text{PO}_4^{3-}$  ratios in the ocean interior are remarkably constant. The specific elemental composition that is the Redfield ratio is truly an 'emergent' property that reflects the interaction of multiple processes, including the acquisition of the elements by plankton, the formation of new biomass and the remineralization of the biomass by bacteria in the ocean interior, as well as losses of nutrients from the ocean because of burial in the sediments (for example, phosphorus in apatite), or outgassing to the atmosphere (for example, production and loss of  $\text{N}_2$ , due to denitrification).

Can the elemental ratios in the ocean change? In principle the answer is 'yes' — but only on timescales comparable to residence times of the nutrients. But why should they

change? They could change as the diversity of organisms in the ocean changes in response to climate or other evolutionary selection mechanisms, or if the bacteria responsible for the remineralization of the organic matter in the oceans shift their preferences in elemental selection. Lakes undergo similar processes as oceans, but their elemental compositions are much more variable. The larger the lake, the less its elemental composition is influenced by the surrounding terrestrial ecosystem, and the more likely it is to track that of the plankton within it.

Although Redfield's concept is 70 years old, it is only just beginning to impact terrestrial ecology. Recent studies on the global patterns of nitrogen and phosphorus in terrestrial plant leaves suggest that these reflect the availability of the two nutrients in soils. But are the nutrients in the soils determined by the N:P ratios in the leaves, or vice versa? The issue of causality — which basically is no longer debated for the oceans — has prompted a search for an analogue for Redfield ratios in terrestrial ecosystems. The search may be a long one, but it will almost certainly result in a better understanding of the global patterns of processes that determine the distribution of the major elements within those ecosystems.

Redfield's concept was an elegant empirical observation that has no simple reductionist explanation, let alone proof. The emergent elemental ratios in the oceans result from scaling the average biochemical composition of organisms in large environments with slow exchanges. As with Newtonian physics, it should not be surprising that the concept fails at very small scales. Nonetheless, it is a powerful organizing principle that illustrates how biological processes at ecosystem levels can alter the distribution of elements on Earth and should be used to help guide us in our understanding of natural biogeochemical patterns and how humans influence them. ■

Paul G. Falkowski is in the Institute of Marine and Coastal Science, Rutgers University, New Jersey 08540, USA. Cabell S. Davis is in the Biology Department, Woods Hole Oceanographic Institution, Massachusetts 02543, USA.

## FURTHER READING

Falkowski, P. G. *et al.* *Science* **305**, 354–360 (2004).

Hedin, L. *Proc. Natl Acad. Sci. USA* **30**, 10849–10850 (2004).

Redfield, A. C. in *James Johnstone Memorial Volume* (ed. Daniel, R. J.) 176–192. (Liverpool Univ. Press, 1934).

Sterner, R. W. & Elser, J. J. *Ecological Stoichiometry: The Biology of the Elements from Molecules to the Biosphere* (Princeton Univ. Press, 2002).



# Into an ice age

Kurt M. Cuffey

Analyses of a new ice core from Greenland yield the first high-resolution picture of the start of the last ice age in the Northern Hemisphere, and of the onset of climate instability as our planet cooled.

**T**he relatively warm and stable climate that humanity has enjoyed for the past 10,000 years will inevitably give way to a new ice age — a tremendous environmental transformation that is destined to bury the sites of Boston, Edinburgh and Stockholm under glacial ice. In *The Day After Tomorrow*, the Hollywood movie most notable for its public abuse of thermodynamics, a new ice age starts in only one week. What does such a transition look like in reality?

A new ice core<sup>1</sup> that samples the entire 3-km thickness of the north-central Greenland ice sheet (Fig. 1) provides us with an unprecedentedly rich and precise view of the onset of the most recent ice age, some 120,000 years ago. And it is from the location of greatest interest — the North Atlantic region, where rapid climate changes have been most dramatic in the past. This achievement, reported on page 147 of this issue<sup>1</sup>, is the result of the efforts of the multinational North Greenland Ice Core Project (NGRIP). Remarkably, individual years of snow deposition are distinguishable for events as far back as 123,000 years in the past — a few years ago, such a resolution was thought to be unattainable.

The story of how the impossible became possible would itself make a fine movie, complete with dramatic scenes of enlightenment in the mass spectrometry lab as isotope analyses reveal past climatic changes. In the early 1990s, two deep ice cores recovered from central Greenland yielded a detailed environmental history extending 100,000 years back in time, through the entirety of the last glacial climate<sup>2</sup>. These showed that climate was frightfully unstable. Within the space of a few decades, the North Atlantic region could evidently warm by 10 °C, while smaller changes of temperature and moisture occurred over wide areas of the planet<sup>3</sup>.

Unfortunately, structural disturbance of the deepest ices<sup>4</sup> prevented these cores from revealing either the onset of the glacial period or events during the preceding warm interglacial period, known as the Eemian, about 130,000 to 120,000 years ago. Initial



**Figure 1** Core territory — the Greenland ice cap, which holds a climate record going back to the last interglacial, the Eemian. Inset, cross-section of an ice core that resulted from drilling slightly offset from the main bore hole; the half-moon shape is composed of Eemian ice.

reports to the contrary<sup>5</sup>, highlighting apparent climate instabilities within the Eemian, were clearly mistaken, as the new report demonstrates<sup>1</sup>. Yet the degree of climate stability during the Eemian is of intense interest: the Eemian was slightly warmer than the world is now, providing an analogue for a possible future climate warmed by atmospheric pollution.

Further to the north of central Greenland (see the map on page 147), the climate is drier. It should therefore be possible to find ice from this last interglacial period safely above the disturbed basal zone. Thus the NGRIP project was born. It seemed likely that this new core would recover a record of the entire last interglacial climate and the start of the ice age. There was little hope for high temporal resolution, however; ice layers are cumulatively thinned as the regional flow compresses them downwards against the subglacial bedrock topography and stretches them horizontally. The vertical compression causing thinning is strongest if there is no melt at the ice sheet bed. And this was precisely the situation expected for the NGRIP site, given that basal temperatures are generally well below the melting point throughout northern and central Greenland.

But the beds of ice sheets, hidden beneath kilometres of opaque ice, contain surprises. The new NGRIP core proved to be located on

something like a volcanic centre — a highly unusual, localized zone of high heat flow from the underlying crust<sup>6</sup>. This heat is melting the basal ice, eliminating the oldest ice layers, and creating a relatively warm and wet environment — which, as the authors point out, merits further investigation for signs of microbial life. This melt also significantly reduces the thinning of deep ice layers, allowing the extraordinary annual resolution of the glacial-onset record reported by the NGRIP team.

The melt has eliminated much of the Eemian ice. Fortunately, however, enough remains to provide a clear record of its latter few millennia, in which the climate was rather stable but warmer than the present by at least 5 °C. The persistence of stable climate under such warm conditions is a key observation. But the loss of the older ice precludes an answer to the even more important question of whether climate was stable in the five or so millennia following the onset of warmth, from about 130,000 years ago, when meltwater runoff from Greenland would probably have been greatest<sup>7</sup>. It has been suggested that increased freshwater flux to the North Atlantic might have the capacity to induce climate change by affecting the ocean's density structure and circulation<sup>8</sup>.

The start of the glacial period was characterized by a mostly gradual cooling, lasting about five millennia<sup>1</sup>. The growth of ice

B. & C. ALEXANDER

WWW.GLACIOLOGYFYKUDK/NGRIP

sheets is necessarily a slow process, limited by the transfer of moisture through the atmosphere, and it appears likely that this process initially limited the rate of climatic cooling. Then, approximately 114,000 years ago, with temperatures having dropped less than halfway to typical full glacial values, the first rapid climate changes began — as documented here for the first time. The timing and characteristics of these events offer an invaluable subject for climate modellers; the mechanisms underlying rapid climate change are still being debated, and climate models have not yet convincingly predicted them.

There is much work yet to be done on the NGRIP core, especially examining the high-resolution characteristics of the record, quantifying the temperature history, and investigating the biogeochemical changes that accompanied the transition to glacial climate. The overview presented in this issue<sup>1</sup> is sufficient to demonstrate that it is a valuable and remarkable core. Yet the NGRIP project has not achieved its primary goal: a reasonably complete record of climate during the last interglacial. How warm did this period get? Were any parts of it climatically unstable? Such information is crucial for evaluating climate models of a warmer world, and for understanding sea-level changes induced by melting of the Greenland

ice sheet. Analysis of basal ices gives direct and compelling evidence that the ice sheet retreated significantly during this period<sup>9</sup>.

There is only one way to fill this gap. A new ice core will have to be extracted from the dry regions of north-central Greenland, but at a safe distance from the heat-flow anomaly discovered at the NGRIP site. The cost and effort of such a project are trivial compared with the possible impact of a rise in sea level, and maybe even rapid climate change, induced by warming of the Arctic region. ■

Kurt M. Cuffey is in the Department of Geography, 507 McCone Hall, University of California, Berkeley, California 94720-4740, USA.

e-mail: [kcuffey@socrates.berkeley.edu](mailto:kcuffey@socrates.berkeley.edu)

1. North Greenland Ice Core Project members *Nature* **431**, 147–151 (2004).
2. Hammer, C., Mayewski, P. A., Peel, D. & Stuiver, M. (eds) *J. Geophys. Res.* **102** (C12), 26317–26886 (1997).
3. Severinghaus, J. P. & Brook, E. J. *Science* **386**, 930–934 (1999).
4. Chappellaz, J., Brook, E., Blunier, T. & Malaize, B. *J. Geophys. Res.* **102**, 26547–26557 (1997).
5. Greenland Ice-Core Project members *Nature* **364**, 203–208 (1993).
6. Fahnstock, M., Abdalati, W., Joughin, I., Brozena, J. & Cugineni, P. *Science* **294**, 2338–2342 (2001).
7. Marshall, S. J. & Cuffey, K. M. *Earth Planet. Sci. Lett.* **179**, 73–90 (2000).
8. Committee on Abrupt Climate Change *Abrupt Climate Change: Inevitable Surprises* (National Academies Press, Washington DC, 2002).
9. Koerner, R. M. *Science* **244**, 964–968 (1989).

**Prokaryotes** Cells lacking a true nucleus. Gene transcription occurs in the cytoplasm.

**Archaeobacteria** Prokaryotes with a plasma membrane of isoprene ether lipids. Protein synthesis occurs on distinctive, archaeobacterial-type ribosomes. Synonymous with Archaea.

**Eubacteria** Prokaryotes with a plasma membrane of fatty acid ester lipids. Protein synthesis occurs on distinctive, eubacterial-type ribosomes. Synonymous with Bacteria.

**Eukaryotes** Cells possessing a true nucleus (lacking in prokaryotes), separated from the cytoplasm by a membrane contiguous with the endoplasmic reticulum (also lacking in prokaryotes). Include double-membrane-bounded cytoplasmic organelles derived from eubacterial endosymbionts<sup>11–13</sup>. The plasma membrane consists of fatty acid ester lipids. Protein synthesis occurs on ribosomes related to the archaeobacterial type. Synonymous with Eucarya.

**Proteobacteria** A name introduced for the group that includes the purple bacteria and relatives<sup>18</sup>. The endosymbiotic ancestor of mitochondria was a member of the proteobacteria as they existed more than 1.4 billion years ago.

Figure 1 Who's who among microbes. In 1938, Edouard Chatton coined the terms prokaryotes and eukaryotes for the organisms that biologists still recognize as such<sup>3</sup>. In 1977 came the report of a deep dichotomy among prokaryotes<sup>19</sup> and designation of the newly discovered groups as eubacteria and archaeobacteria. In 1990, it was proposed<sup>2</sup> to rename the eukaryotes, eubacteria and archaeobacteria as eucarya, bacteria and archaea. Although widely used, the latter names left the memberships of these groups unchanged, so the older terms have priority.

## Evolutionary biology

# Early evolution comes full circle

William Martin and T. Martin Embley

Biologists use phylogenetic trees to depict the history of life. But according to a new and roundabout view, such trees are not the best way to summarize life's deepest evolutionary relationships.

Charles Darwin described the evolutionary process in terms of trees, with natural variation producing diversity among progeny and natural selection shaping that diversity along a series of branches over time. But in the microbial world things are different, and various schemes have been devised to take both traditional and molecular approaches to microbial evolution into account. Rivera and Lake (page 152 of this issue<sup>1</sup>) provide the latest such scheme, based on analysing whole-genome sequences, and they call for a radical departure from conventional thinking.

Unknown to Darwin, microbes use two mechanisms of natural variation that disobey the rules of tree-like evolution: lateral gene transfer and endosymbiosis. Lateral gene transfer involves the passage of genes among distantly related groups, causing branches in the tree of life to exchange bits of their fabric. Endosymbiosis — one cell living within another — gave rise to the double-membrane-bounded organelles of

eukaryotic cells: mitochondria (the powerhouses of the cell) and chloroplasts (of no further importance here). At the endosymbiotic origin of mitochondria, a free-living proteobacterium came to reside within an archaeobacterially related host — see Fig. 1 for terminology. This event involved the genetic union of two highly divergent cell lineages, causing two deep branches in the tree of life to merge outright. To this day, biologists cannot agree on how often lateral gene transfer and endosymbiosis have occurred in life's history; how significant either is for genome evolution; or how to deal with them mathematically in the process of reconstructing evolutionary trees. The report by Rivera and Lake<sup>1</sup> bears on all three issues. And instead of a tree linking life's three deepest branches (eubacteria, archaeobacteria and eukaryotes), they uncover a ring.

The ring comes to rest on evolution's sorest spot — the origin of eukaryotes. Biologists fiercely debate the relationships between eukaryotes (complex cells that have a nucleus

and organelles) and prokaryotes (cells that lack both). For a decade, the dominant approach has involved another intracellular structure called the ribosome, which consists of complexes of RNA and protein, and is present in all living organisms. The genes encoding an organism's ribosomal RNA (rRNA) are sequenced, and the results compared with those for rRNAs from other organisms. The ensuing tree<sup>2</sup> divides life into three groups called domains (Fig. 2a). The usefulness of rRNA in exploring biodiversity within the three domains is unparalleled, but the proposal for a natural system of all life based on rRNA alone has come increasingly under fire.

Ernst Mayr<sup>3</sup>, for example, argued forcefully that the rRNA tree errs by showing eukaryotes as sisters to archaeobacteria, thereby obscuring the obvious natural division between eukaryotes and prokaryotes at the level of cell organization (Fig. 2b). A central concept here is that of a tree's 'root', which defines its most ancient branch and hence the relationships among the deepest-diverging



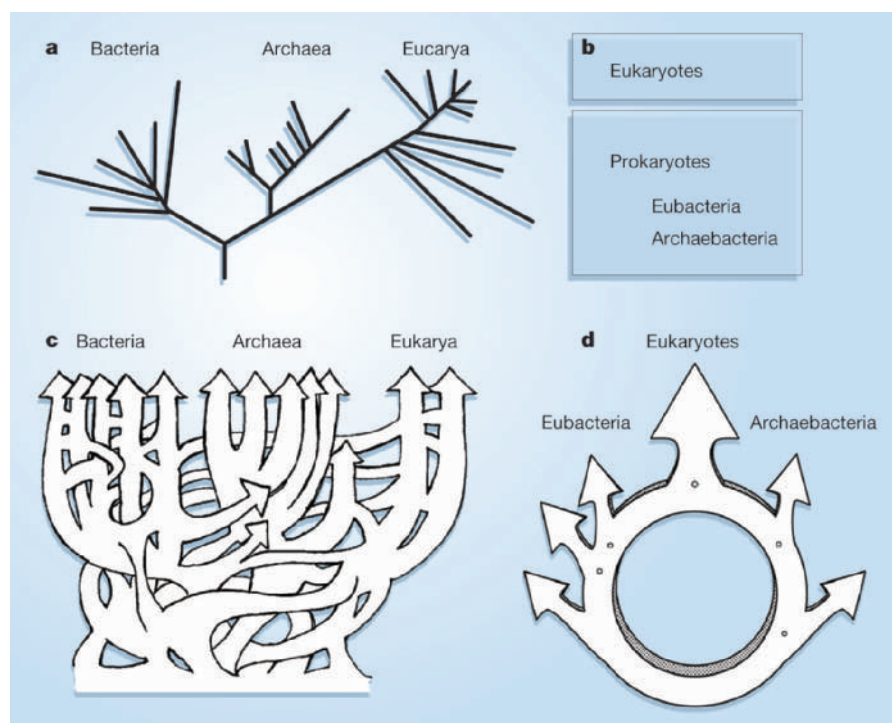
lineages. The eukaryote–archaeobacteria sister-grouping in the rRNA tree hinges on the position of the root (the short vertical line at the bottom of Fig. 2a). The root was placed on the eubacterial branch of the rRNA tree based on phylogenetic studies of genes that were duplicated in the common ancestor of all life<sup>2</sup>. But the studies that advocated this placement of the root on the rRNA tree used, by today's standards, overly simple mathematical models and lacked rigorous tests for alternative positions<sup>4</sup>.

One discrepancy is already apparent in analyses of a key data set used to place the root, an ancient pair of related proteins, called elongation factors, that are essential for protein synthesis<sup>5</sup>. Although this data set places the root on the eubacterial branch, it also places eukaryotes within the archaeobacteria, not as their sisters<sup>5</sup>. Given the uncertainties of deep phylogenetic trees based on single genes<sup>4</sup>, a more realistic view is that we still don't know where the root on the rRNA tree lies and how its deeper branches should be connected.

A different problem with the rRNA tree, as Ford Doolittle<sup>6</sup> has argued, is that lateral gene transfer pervades prokaryotic evolution. In that view, there is no single tree of genomes to begin with, and the concept of a natural system with bifurcating genome lineages should be abandoned (Fig. 2c). Added to that are genome-wide sequence comparisons showing eukaryotes to possess far more eubacteria-like genes than archaeobacteria-like genes<sup>7,8</sup>, in diametric opposition to the rooted rRNA tree, which accounts for only one gene. Despite much dissent, the rRNA tree has nonetheless dominated biologists' thinking on early evolution because of the lack of better alternatives.

Rivera and Lake's ring of life<sup>1</sup> (Fig. 2d) includes the analysis of hundreds of genes, not just one. It puts prokaryotes in one bin and eukaryotes in another<sup>3</sup>; it allows lateral gene transfer to be used in assessing genome-based phylogeny<sup>7</sup>; and it recovers the connections between prokaryote and eukaryote genomes as no single gene tree possibly could. Their method — 'conditioned reconstruction' — uses shared genes as a measure of genome similarity but does not discriminate between vertically or horizontally inherited genes. This method does not uncover all lateral gene transfer in all genomes. But it does uncover the dual nature of eukaryotic genomes<sup>7,8</sup>, which in the new scheme sit simultaneously on a eubacterial branch and an archaeobacterial branch. This is what seals the ring.

As the simplest interpretation of the ring, Rivera and Lake<sup>1</sup> propose that eukaryotic chromosomes arose from a union of archaeobacterial and eubacterial genomes. They suggest that the biological mechanism behind that union was an endosymbiotic association between two prokaryotes. The ring is thus at



**Figure 2** Four schemes of natural order in the microbial world. **a**, The three-domain proposal based on the ribosomal RNA tree, as rooted with data from anciently duplicated protein genes. **b**, The two-empire proposal, separating eukaryotes from prokaryotes and eubacteria from archaeobacteria. **c**, The three-domain proposal, with continuous lateral gene transfer among domains. **d**, The ring of life, incorporating lateral gene transfer but preserving the prokaryote–eukaryote divide. (Redrawn from refs 2, 3, 6 and 1, respectively.)

odds with the view of eukaryote origins by simple Darwinian divergence<sup>9,10</sup>, but is consistent with symbiotic models of eukaryote origins, variants of which abound<sup>11</sup>. Some symbiotic models suggest that an archaeobacterium–eubacterium symbiosis was followed by the endosymbiotic origin of mitochondria; others suggest that the host cell in which mitochondria settled was an archaeobacterium outright.

Rivera and Lake's findings do not reveal whether a symbiotic event preceded the mitochondrion. But — importantly — they cannot reject the mitochondrial endosymbiont as the source of the eubacterial genes in eukaryotes. The persistence of the mitochondrial compartment, especially in anaerobic eukaryotic lineages<sup>12,13</sup>, among which the most ancient eukaryote lineages have traditionally been sought, provides phylogeny-independent evidence that the endosymbiotic origin of mitochondria occurred in the eukaryotic common ancestor. Phylogeny-independent evidence for any earlier symbiosis is lacking. So the simpler, hence preferable, null hypothesis is that eubacterial genes in eukaryotes stem from the mitochondrial endosymbiont.

Rejecting that null hypothesis will require improved mathematical tools for probing deep phylogeny. Indeed, it is not clear if conditioned reconstruction alone is sensitive enough to do this — analyses of individual genes are still needed. But

eukaryotes are more than 1.4 billion years old<sup>14</sup> and such time-spans push current tree-building methods to, and perhaps well beyond, their limits<sup>15</sup>.

Looking into the past with genes is like gazing at the stars with telescopes: it involves a lot of mathematics<sup>16</sup>, most of which the stargazers never see. With better telescopes we can see more details further back in time, but nobody knows for sure how good today's gene-telescopes really are. Mathematicians have a well-developed theory for building trees from recently diverged gene sequences<sup>17</sup>, but mathematical methods for recovering ancient mergers in the history of life are still rare. Rivera and Lake's ring depicts the eukaryotic genome for what it is — a mix of genes with archaeobacterial and eubacterial origins.

William Martin is at the Institut für Botanik III, Heinrich-Heine Universität Düsseldorf, 40225 Düsseldorf, Germany.  
e-mail: w.martin@uni-duesseldorf.de  
T. Martin Embley is in the School of Biology, The Devonshire Building, University of Newcastle upon Tyne, Newcastle upon Tyne NE1 7RU, UK.  
e-mail: martin.embly@ncl.ac.uk

1. Rivera, M. C. & Lake, J. A. *Nature* **431**, 152–155 (2004).
2. Woese, C., Kandler, O. & Wheelis, M. L. *Proc. Natl Acad. Sci. USA* **87**, 4576–4579 (1990).
3. Mayr, E. *Proc. Natl Acad. Sci. USA* **95**, 9720–9723 (1998).
4. Penny, D., Hendy, M. D. & Steel, M. A. in *Phylogenetic Analysis of DNA Sequences* (eds Miyamoto, M. M. & Cracraft, J.) 155–183 (Oxford Univ. Press, 1991).
5. Baldauf, S., Palmer, J. D. & Doolittle, W. F. *Proc. Natl Acad. Sci. USA* **93**, 7749–7754 (1996).

6. Doolittle, W. F. *Science* **284**, 2124–2128 (1999).
7. Rivera, M. C., Jain, R., Moore, J. E. & Lake, J. A. *Proc. Natl Acad. Sci. USA* **95**, 6239–6244 (1998).
8. Esser, C. et al. *Mol. Biol. Evol.* **21**, 1643–1660 (2004).
9. Kandler, O. in *Early Life on Earth* (ed. Bengtson, S.) 152–160 (Columbia Univ. Press, New York, 1994).
10. Woese, C. R. *Proc. Natl Acad. Sci. USA* **99**, 8742–8747 (2002).
11. Martin, W., Hoffmeister, M., Rotte, C. & Henze, K. *Biol. Chem.* **382**, 1521–1539 (2001).
12. Embley, T. M. et al. *IUBMB Life* **55**, 387–395 (2003).
13. Tovar, J. et al. *Nature* **426**, 172–176 (2003).
14. Javaux, E. J., Knoll, A. H. & Walter, M. R. *Nature* **412**, 66–69 (2001).
15. Penny, D., McComish, B. J., Charleston, M. A. & Hendy, M. D. *J. Mol. Evol.* **53**, 711–723 (2001).
16. Semple, C. & Steel, M. A. *Phylogenetics* (Oxford Univ. Press, 2003).
17. Felsenstein, J. *Inferring Phylogenies* (Sinauer, Sunderland, MA, 2004).
18. Stackebrandt, E., Murray, R. G. E. & Trüper, H. G. *Int. J. Syst. Bact.* **38**, 321–325 (1988).
19. Woese, C. R. & Fox, G. E. *Proc. Natl Acad. Sci. USA* **74**, 5088–5090 (1977).

## Neurobiology

## Feeding the brain

Claire Peppiatt and David Attwell

In computationally active areas of the brain, the blood flow is increased to provide more energy to nerve cells. New data fuel the controversy over how this energy supply is regulated.

Like all tissues, our brains need energy to function, and this comes in the form of oxygen and glucose, carried in the blood. The brain's information-processing capacity is limited by the amount of energy available<sup>1</sup>, so, as has been recognized for more than a century, blood flow is increased to brain areas where nerve cells are active<sup>2</sup>. This increase in flow provides the basis for functional magnetic resonance imaging of brain activity<sup>2</sup>, but exactly how the flow is increased is uncertain. On page 195 of this issue, Mulligan and MacVicar<sup>3</sup> reveal a previously unknown role for non-neuronal brain cells called astrocytes in controlling the brain's blood flow. Intriguingly, the new data contradict a previous suggestion for how astrocytes regulate flow.

Figure 1 shows recent developments in our understanding of how the blood flow in the brain is controlled. Glucose and oxygen are provided to neurons through the walls of capillaries, the blood flow through which is controlled by the smooth muscle surrounding precapillary arterioles. Dedicated neuronal networks in the brain signal to the smooth muscle to constrict or dilate arterioles and thus decrease or increase blood flow<sup>2</sup>; for example, neurons that release the neurotransmitter molecule noradrenaline constrict arterioles. In addition, the neuronal activity associated with information processing increases local blood flow. This is in part due to neurons that release the transmitter glutamate, which raises the intracellular concentration of  $\text{Ca}^{2+}$  ions in other neurons, thereby activating the enzyme nitric oxide (NO) synthase and leading to the release of NO. This in turn dilates arterioles<sup>4</sup>.

A radical addition to this scheme came with the claim of Zonta et al.<sup>5</sup> that glutamate also works through astrocytes in the brain to dilate arterioles. Glutamate raises the  $\text{Ca}^{2+}$  concentration in astrocytes, and thus activates the enzyme phospholipase  $\text{A}_2$ , which produces a fatty acid, arachidonic acid. This

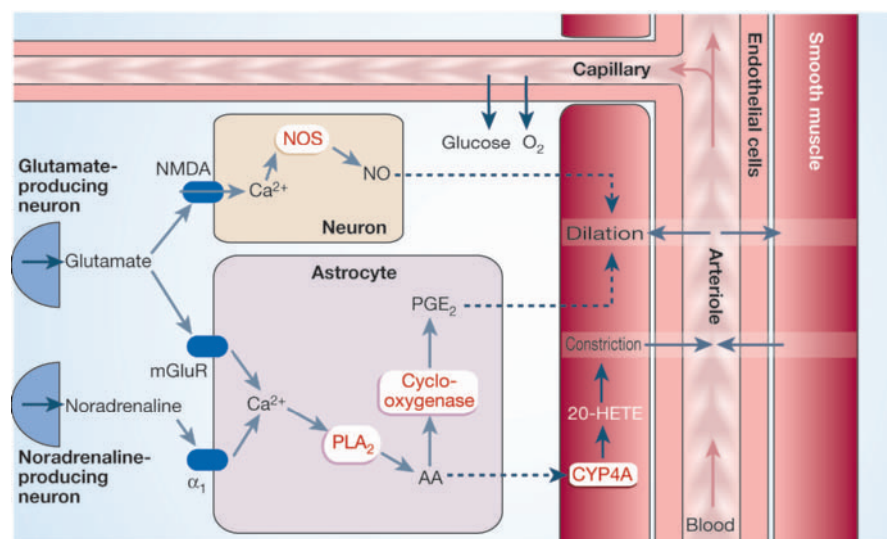
is converted by the enzyme cyclooxygenase into prostaglandin derivatives, which dilate arterioles. An attractive aspect of a role for astrocytes in controlling blood flow is that, although most of their cell membrane surrounds neurons and so can sense neuronal glutamate release, they also send out an extension, called an endfoot, close to blood vessels: thus, astrocyte anatomy is ideal for regulating blood flow in response to local neuronal activity<sup>6</sup>. In this scheme, a rise in the  $\text{Ca}^{2+}$  levels in astrocytes, just like in neurons, would dilate arterioles and increase local blood flow.

The new data contradict these results. Mulligan and MacVicar<sup>3</sup> inserted a 'caged' form of  $\text{Ca}^{2+}$  into astrocytes in brain slices

taken from rats and mice. By using light to suddenly uncage the  $\text{Ca}^{2+}$ , they found that an increase in the available  $\text{Ca}^{2+}$  concentration within astrocytes produces a constriction of nearby arterioles that could powerfully decrease local blood flow (the 23% decrease in diameter seen would increase the local resistance to blood flow threefold, by Poiseuille's law).

They show that this constriction results from  $\text{Ca}^{2+}$  activating phospholipase  $\text{A}_2$  to generate arachidonic acid, as above; the twist is that this arachidonic acid is then processed by a cytochrome P450 enzyme (CYP) into a constricting derivative. The authors propose that this derivative is 20-hydroxyeicosatetraenoic acid (20-HETE), formed by CYP4A in the arteriole smooth muscle<sup>7</sup> (but the high concentration of CYP4A blocker used to deduce this might also block other enzymes<sup>8</sup>). The authors also found that noradrenaline evoked a rise in astrocyte  $\text{Ca}^{2+}$  concentration and arteriole constriction. Unexpectedly, therefore, it seems that rather than noradrenaline-producing neurons signalling directly to smooth muscle, as is conventionally assumed, much of their constricting action may be mediated indirectly by astrocytes. In fact this is consistent with the finding that many noradrenaline-release sites on neurons are located near astrocytes<sup>9</sup>.

Is it possible to reconcile the new data<sup>3</sup> (a rise in astrocyte  $\text{Ca}^{2+}$  levels constricts arterioles) with those of Zonta et al.<sup>5</sup> (a rise in  $\text{Ca}^{2+}$  dilates arterioles)? A likely solution is that the increased concentration of  $\text{Ca}^{2+}$  in astrocytes leads to the production of both constricting



**Figure 1 Controlling blood flow in the brain.** Computationally active neurons release glutamate (top left). This activates neuronal NMDA-type receptors,  $\text{Ca}^{2+}$  influx through which leads to nitric oxide synthase (NOS) releasing NO, which works on smooth muscle to dilate arterioles. This increases the supply of oxygen and glucose to the brain. Glutamate also spills over to astrocyte receptors (mGluRs), which raise the  $\text{Ca}^{2+}$  levels in astrocytes and generate arachidonic acid (AA) via phospholipase  $\text{A}_2$  (PLA<sub>2</sub>). Cyclooxygenase-generated derivatives of AA (PGE<sub>2</sub>) dilate arterioles<sup>5</sup>, whereas, as Mulligan and MacVicar show<sup>3</sup>, the CYP4A-generated derivative 20-HETE constricts them. Astrocyte  $\text{Ca}^{2+}$  levels can also be raised by noradrenaline — released from dedicated neurons that control the circulation — which works through  $\alpha_1$  receptors (bottom left). Dotted lines show messengers diffusing between cells. The detailed anatomy of synapses and astrocytes is not portrayed.



CYP-generated derivatives of arachidonic acid and dilating cyclooxygenase-generated derivatives. The relative importance of these opposing effects may have differed in the two groups' experiments because of a difference in the age of the animals or the brain area studied: Mulligan and MacVicar looked at the hippocampus of 13–18-day-old rodents; Zonta *et al.* focused on the cerebral cortex at 9–15 days. (Furthermore, outside the brain, whether 20-HETE constricts or dilates depends on both tissue area and age<sup>10</sup>.)

The relative importance of the constricting and dilating signals will also depend on the initial baseline level of contraction of the arteriole smooth muscle. Mulligan and MacVicar found that inhibiting NO synthase converted the astrocyte-evoked constriction into a dilation, implying that the dilation observed by Zonta *et al.* might be unphysiological and due to their inhibiting NO synthase. However, a rationale for blocking this enzyme is that, *in vivo*, blood flow generates a baseline level of contraction of arteriole smooth muscle, which is normally absent in brain-slice experiments but can be mimicked by blocking the dilating effects of NO released from neurons and blood vessels. The change from constriction to dilation that Mulligan and MacVicar saw when NO synthase was blocked may reflect a stronger, perhaps more physiological, prior baseline contraction of arteriole smooth muscle in the absence of NO, because from a more constricted baseline it is harder to evoke further constriction and easier to evoke dilation.

Finally, another explanation for the difference between the two groups' results could be that, whereas Mulligan and MacVicar's  $\text{Ca}^{2+}$ -uncaging experiments elegantly isolate astrocyte-initiated signalling, Zonta and colleagues' use of a glutamate analogue to raise the astrocyte  $\text{Ca}^{2+}$  concentration could also activate non-astrocyte signalling pathways, for example in the endothelial cells that line blood vessels<sup>11</sup>. These pathways might also be activated by glutamate released *in vivo*.

Excitingly, Mulligan and MacVicar's discovery of an astrocyte-mediated constriction pathway may also give insight into how arterioles are dilated, in both normal and pathological circumstances. Although normal dilation occurs partly through NO increasing the levels of the molecule cyclic GMP in smooth muscle, NO also inhibits the production of 20-HETE<sup>7</sup>. So part of NO's dilating effects may result from a suppression of the astrocyte-initiated signalling discovered by Mulligan and MacVicar. The production of 20-HETE is also inhibited when oxygen levels fall<sup>7</sup>, possibly contributing to the dilation produced by low oxygen levels in the brain.

Eventual reconciliation of the results of Zonta *et al.*<sup>5</sup> and Mulligan and MacVicar<sup>3</sup> may require difficult experiments involving perfusing arterioles in brain slices with an

artificial extracellular fluid, to generate a physiological level of baseline contraction<sup>12</sup>. Nonetheless, it is clear that astrocytes play a starring role in the control of blood flow in the brain.

Claire Peppiatt and David Attwell are in the Department of Physiology, University College London, Gower Street, London WC1E 6BT, UK. e-mail: d.attwell@ucl.ac.uk

- Laughlin, S. B. & Sejnowski, T. J. *Science* **301**, 1870–1874 (2003).
- Attwell, D. & Iadecola, C. *Trends Neurosci.* **25**, 621–625 (2002).

- Mulligan, S. J. & MacVicar, B. A. *Nature* **431**, 195–199 (2004).
- Faraci, F. M. & Breese, K. R. *Circ. Res.* **72**, 476–480 (1993).
- Zonta, M. *et al.* *Nature Neurosci.* **6**, 43–50 (2003).
- Simard, M. *et al.* *J. Neurosci.* **23**, 9254–9262 (2003).
- Roman, R. J. *Physiol. Rev.* **82**, 131–185 (2002).
- Miyata, N. *et al.* *Br. J. Pharmacol.* **133**, 325–329 (2001).
- Cohen, Z., Molinatti, G. & Hamel, E. *J. Cereb. Blood. Flow Metab.* **17**, 894–904 (1997).
- Fuloria, M., Eckman, D. M., Leach, D. A. & Aschner, J. L. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **287**, L360–L365 (2004).
- Gillard, S. E., Tzaferis, J., Tsui, H. C. & Kingston, A. E. *J. Comp. Neurol.* **461**, 317–332 (2003).
- Brown, L. A., Key, B. J. & Lovick, T. A. *Pflügers Arch.* **443S**, S260 (2002).

## Condensed-matter physics

# The qubit and the cavity

Yu. Makhlin, G. Schön and A. Shnirman

Coherent coupling of superconducting qubits and electromagnetic modes, analogous to atom–photon coupling, has been demonstrated — a step towards the communication of quantum information.

For some time, the efforts of the quantum-optics community have been dedicated to reaching the so-called strong-coupling regime of cavity quantum electrodynamics (QED)<sup>1</sup>. In cavity QED, atoms are strongly coupled to photons. Usually, photons propagate freely in space over large distances and interact with atoms only briefly as they pass by. The interaction is stronger if photons and atoms are trapped together inside an optical cavity — this consists of a region bounded by two or more mirrors that are aligned to provide multiple reflections of light waves. Suppose that an atom in an excited state enters an empty cavity; it can repeatedly and coherently emit and reabsorb a photon, in a cycle called vacuum Rabi oscillation. In the desired regime, many oscillations occur before the photon or atom escape from the cavity. Starting on page 159 of this issue, Wallraff *et al.*<sup>2</sup> and Chiorescu *et al.*<sup>3</sup> report experiments in which the concept of cavity QED has been realized in a solid-state system, using superconducting Josephson junctions. These experiments constitute a major step towards quantum communication with solid-state systems.

Josephson qubits (quantum bits) are effectively artificial atoms — two-level quantum systems that use either electric charge or magnetic flux as the primary quantum degree of freedom. They are based on Josephson junctions, which are sandwiches of two layers of superconducting material separated by an insulating layer. Progress in experiments with Josephson qubits has included the demonstration of coherent manipulations, as well as free and driven oscillations, in single<sup>4–6</sup> and coupled<sup>7</sup> qubits. Now Wallraff *et al.*<sup>2</sup>, inspired by quantum optical cavity QED, have coupled a Josephson qubit (the 'atom'), on a chip, to a harmonic-oscillator mode of a transmission line (the superconducting 'cavity'). In atomic

systems, care must be taken to extend the time the atoms spend in a cavity, but here the advantage is that the Josephson qubit is fixed on the chip. The qubit is positioned at a point of maximum amplitude for a standing wave in the cavity, and this induces a strong dipole coupling between the charge of the qubit and the electric field of the wave.

Wallraff *et al.*<sup>2</sup> investigated the strong-coupling regime using spectroscopy. The qubit–wave coupling causes two excited energy states in the system to split; electromagnetic radiation sent into the cavity can probe the structure of the energy states by inducing transitions between the ground state and the two split states (Fig. 1). The energy gap between the split states sets the frequency of the vacuum Rabi oscillations. Wallraff *et al.* recorded a splitting energy equivalent to a frequency of about 10 MHz — much higher than the widths of the levels, typically 1 MHz, that characterize the timescale on which the system loses coherence through dissipation. This means that about ten vacuum Rabi oscillations could be observed in a time-resolved experiment.

Chiorescu *et al.*<sup>3</sup> have studied the coupling of a magnetic-flux Josephson qubit to a different kind of oscillation — the plasma oscillations in a SQUID (a superconducting quantum interference device, which comprises two Josephson junctions, on either side of a loop of wire). Chiorescu *et al.* proved that coherent coupling was occurring between the qubit and the oscillator using Rabi oscillations of various kinds. First, they induced Rabi oscillations in the qubit: by shining a resonant microwave signal onto the qubit, a regime can be reached in which the microwave mode periodically transfers the energy of one photon to the qubit and then gains it back. In contrast to the vacuum Rabi oscillations, this process is induced by a classical electromagnetic field involving many photons.

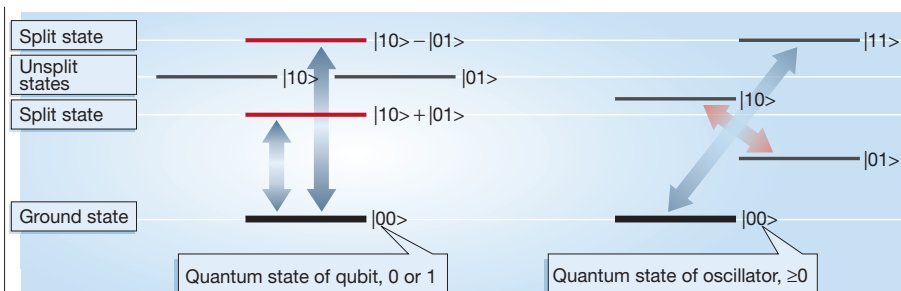


Figure 1 The structure of the energy levels in a coupled qubit–oscillator system, as studied by Wallraff *et al.*<sup>2</sup> (left) and Chiorescu *et al.*<sup>3</sup> (right).

As a further test, Chiorescu *et al.* induced and then analysed coherent Rabi oscillations between the energy levels of the coupled qubit–oscillator system; again, their data are in agreement with expectation. When they prepared the qubit in the ground state, only the Rabi processes associated with this state were visible; similarly, when the qubit was prepared in an excited state, only the Rabi oscillations involving that state were observed (Fig. 1). Although the coherence time was, at most, tens of nanoseconds, the demonstration of time-resolved coherent oscillations in this combined system is a remarkable achievement. The measured frequency of these processes, the Rabi frequency, passed the litmus test — it depended linearly on the amplitude of the microwave that induced the oscillations. Moreover, as the microwave power, and hence the Rabi frequency, was varied, the authors saw a resonance between the Rabi oscillations and the SQUID plasma oscillations, at exactly the position expected.

These experiments<sup>2,3</sup> show that a high level of coherence and control is possible in Josephson circuits, such that, on a single chip, qubits could be coupled through the oscillator to create a quantum computer (in the spirit of the coupling in ion-trap quantum registers<sup>8</sup>). This is also a step closer to

quantum communication — the transfer of quantum information between relatively distant stationary qubits via propagating waves. In both experiments, there is circumstantial evidence that the qubit and the oscillator were entangled; in future work with electronic qubit–oscillator or multi-qubit<sup>7</sup> circuits, it should be possible to observe the degree of entanglement directly. However, this would require the ability to carry out ‘single-shot’ measurements of both subsystems: although detectors are available that come close to achieving this<sup>5,6,9</sup>, combining all the ingredients into a single circuit and operating it in a coherent regime is a challenge still to be faced.

Yu. Makhlin, G. Schön and A. Shnirman are at the Institut für Theoretische Festkörperphysik, University of Karlsruhe, D-76128 Karlsruhe, Germany.

e-mail: schoen@tfp.physik.uni-karlsruhe.de

1. Mabuchi, H. & Doherty, A. *Science* **298**, 1372–1377 (2002).
2. Wallraff, A. *et al.* *Nature* **431**, 162–167 (2004).
3. Chiorescu, I. *et al.* *Nature* **431**, 159–162 (2004).
4. Nakamura, Y., Pashkin, Yu. A. & Tsai, J. S. *Nature* **398**, 786–788 (1999).
5. Vion, D. *et al.* *Science* **296**, 886–889 (2002).
6. Chiorescu, I. *et al.* *Science* **299**, 1869–1871 (2003).
7. Pashkin, Yu. A. *et al.* *Nature* **421**, 823–826 (2003).
8. Cirac, J. I. & Zoller, P. *Phys. Rev. Lett.* **74**, 4091–4094 (1995).
9. Astafiev, O., Pashkin, Yu. A., Yamamoto, T., Nakamura, Y. & Tsai, J. S. *Phys. Rev. B* **69**, 180507 (2004).

## Animal behaviour

# Relative size in the mating game

Malte Andersson and Johan Wallander

A common trend in size differences between males and females is a long-standing puzzle. A study of shorebirds shows that the type and strength of competition for mates may explain much of the pattern.

Five decades ago, the German evolutionary biologist Bernhard Rensch<sup>1</sup> presented an intriguing rule for the differing sizes of male and female animals. He found that in several groups (clades) that each contain related species, male size relative to female size increases with the body size of the species. Rensch’s rule has since been verified in animals as diverse as arthropods, reptiles, birds and mammals, including primates<sup>2</sup>. The causes behind the rule, however,

have remained unclear<sup>2–5</sup>. Why are males much larger than females in many animals with large body size? And why, in the same clade, are males similar or even smaller than females in species with small body size<sup>1,2</sup>?

As they describe in *Proceedings of the National Academy of Sciences*, Székely *et al.*<sup>6</sup> have carried out a comparative analysis of shorebirds that show such trends, and have come up with some thought-provoking conclusions. They show that the trend in sexual



## 100 YEARS AGO

Are they not methodologically equivalent, the three systems of classification — (a) of plants into herbs, shrubs and trees; (b) of animals into birds, beasts and fishes; and (c) of humans into the sanguine, the lymphatic, the bilious and the melancholy? Why, then, is it that science, having long ago given us a *Systema Naturae* and a *nomenclature botanicus and zoologicus*, still leaves us almost without the rudiments of a *Systema Hominis* and a *nomenclature sociologicus*? It may be asked in reply, What of the anthropologists and their half-century of taxonomic labours in the name of science? But the anthropological classifications belong, in appearance at least, to natural and not human history. They do not rise through psychology into sociology... Of late the anthropologist has shown signs of attaching himself to the psychologist; and this suggests another form of the initial question, Why have anthropologists not endeavoured to formulate even a provisional classification of psychological types? Why have they, with unconscious naïvete, been content to accept implicitly the popular classification that traditionally survives from early Greek thought? To this question the positivist will be ready with his answer, but perhaps it were wiser to leave it as a shameful reminder to the laggard sociologist.

From *Nature* 8 September 1904.

## 50 YEARS AGO

The passing of Prof. T. F. Dreyer has deprived the study of early man in South Africa of one of its acknowledged leaders, and his place will not be easily filled. Outside South Africa, Prof. Dreyer will be most widely remembered as the discoverer of the Florisbad skull, the most remarkable human fossil to be found in Africa since the Broken Hill skull. This discovery was a well-deserved reward for his intuition in selecting for thorough investigation the Florisbad mineral spring deposits with their wealth of archaeological and fossil mammalian remains. But his explorations in the Matjes River Cave and elsewhere also constitute notable contributions to our knowledge of man in South Africa from prehistoric to historic times. With a characteristic scorn for the compartmenting of knowledge, Prof. Dreyer pursued his studies simultaneously in the field of physical anthropology, Quaternary mammalian palaeontology, archaeology and even Quaternary geology and climatology.

From *Nature* 11 September 1954.



size difference (SSD) can be explained by two aspects of sexual selection (which arises from competition over mates), and the interaction between them. One aspect is the strength of sexual selection involved; the other is the agility of the male's display. The outcome of the analysis is similar whether it is based on body mass or wing length as a measure of size.

Males of many animals compete in fights for mates, and such contests favour large body size (Fig. 1). Through genetic correlations between the sexes<sup>2,4,5</sup>, such competition may also lead to some increase in female body size, though less so than in males. In consequence, the mean size of each sex will increase, and so will the relative size difference between them. Increased male-biased SSD therefore tends to become associated with large body size. On the other hand, there are several smaller species in which sex roles are reversed and females compete strongly for males. Such species tend to have female-biased SSDs<sup>5</sup>.

Some forms of male competition can favour smaller males. For example, in species where males compete by acrobatic aerial displays, there may be strong sexual selection for small male body size. For geometrically similar animals, agility increases with reduced body size<sup>7</sup>, and this might lead to higher mating success of smaller males in certain birds<sup>8,9</sup>. Through genetic correlations it may also lead to some reduction in female body size. This in turn can help to explain why female-biased SSD increases with reduced body size in some birds and other animals with agile male display<sup>5</sup>.

Székely *et al.*<sup>6</sup> tested the relative role of these mechanisms in producing SSD among 102 species of shorebirds and their allies (a clade named Charadriides). This is an ideal group for the purpose as it encompasses great variation in the traits of interest<sup>6,8,10</sup>. The analyses confirm Rensch's rule, showing that males are bigger than females in most large species, and that male-biased SSD increases with body size. In contrast, females are bigger than males in many small species.

The social mating system of a species can be used as a proxy for the strength of sexual selection in males. The strength increases from polyandry (where females compete to have several mates), through monogamy, to polygyny (where males compete to have several mates). As predicted, Székely *et al.* find that stronger sexual selection is associated with increasing male-biased SSD, and more agile aerial display is associated with increasing female-biased SSD. Notably, there are strong interactions between agility and strength of sexual selection. In polygynous species, the female is larger if male displays involve agility (Fig. 2); if they don't, males are the larger sex. Forms with less male competition (monogamy or polyandry) tend to have larger females regardless of male agility (Fig. 2). Such interactions may help to explain



Figure 1 Competition at a mating arena. Males are the larger sex in many groups of animals in which males fight over females. Ruffs (*Philomachus pugnax*), for example, are shorebirds in which there is a great size difference between the sexes. Here, one of two (larger, collared) males mates with one of two females.

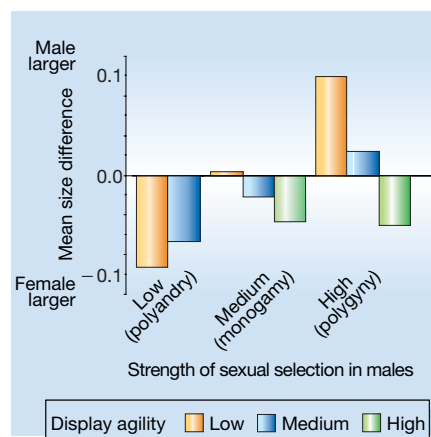


Figure 2 Székely and colleagues' analysis<sup>6</sup> shows that the size difference between male and female shorebirds depends on the type as well as the strength of sexual selection. In species with strong male competition (polygyny) and non-agile display, males are larger than females. In species with agile aerial display, males are smaller than females. When competition over mates is weaker in males and strong in females (polyandry), females are the larger sex. The results are based on 75 species of shorebirds, with size differences measured as the mean of  $\log_{10}(\text{male mass}) - \log_{10}(\text{female mass})$ . (Modified after ref. 6.)

why some groups with intense male competition show a range of SSD patterns, from female bias to male bias.

Sex difference in choice of habitat or food might also influence SSD. Székely *et al.* find that sexual difference in bill length — a trait of crucial importance for foraging — does not follow Rensch's rule. And when male and female body sizes are controlled statistically, sex difference in bill length is not associated with display agility or strength of sexual selection. So, unlike body size, which depends strongly on sexual selection, SSD in bill length is influenced mainly by other (probably ecological) selection pressures.

The generality of these conclusions can be explored by deriving predictions from them for tests in other animals. One prediction is that large males are favoured in species with large body size that compete for females in contests requiring strength or endurance<sup>5</sup> rather than agility. Moreover, among small species in the same clades, sex-role reversal and stronger competition among females, or male competition by agile<sup>8</sup> and energy-demanding<sup>5</sup> display, are expected to favour reduced male size, sometimes leading to female-biased SSD.

A review<sup>2</sup> of SSD in 40 groups shows that exceptions to Rensch's rule are rare and mostly occur in animals where females are the larger sex. There is much scope for testing the predictions in these and other groups. Advances in comparative and other analyses of SSD<sup>2,6</sup> open possibilities for clarifying the striking patterns in sexual size difference that have long puzzled biologists. Analyses of the form and strength of sexual selection are likely to be crucial in such work, as shown by Székely and colleagues' study of shorebirds. There is plenty of interesting work ahead — for instance, it is still not clear why, in most clades where some species show stronger sexual selection in females than males, this occurs mainly in its smaller forms.

Malte Andersson and Johan Wallander are in the Department of Zoology, University of Gothenburg, SE 405 30 Göteborg, Sweden.

e-mail: malte.andersson@zoool.gu.se

1. Rensch, B. *Evolution Above the Species Level* (Columbia Univ. Press, New York, 1959).
2. Fairbairn, D. J. *Annu. Rev. Ecol. Syst.* **28**, 659–687 (1997).
3. Clutton-Brock, T. H. *et al. Nature* **269**, 797–800 (1977).
4. Maynard Smith, J. *The Evolution of Sex* (Cambridge Univ. Press, 1978).
5. Andersson, M. *Sexual Selection* (Princeton Univ. Press, 1994).
6. Székely, T., Freckleton, R. P. & Reynolds, J. D. *Proc. Natl. Acad. Sci. USA* **101**, 12224–12227 (2004).
7. Andersson, M. & Norberg, R. Å. *Biol. J. Linn. Soc.* **15**, 105–130 (1981).
8. Jehl, J. R. & Murray, B. G. *Curr. Ornithol.* **6**, 1–86 (1986).
9. Blomqvist, D. *et al. Anim. Behav.* **54**, 1291–1299 (1997).
10. Székely, T., Reynolds, J. D. & Figuerola, J. *Evolution* **54**, 1404–1413 (2000).

# An unexpected social servant

Christine Le Roy and Jeffrey L. Wrana

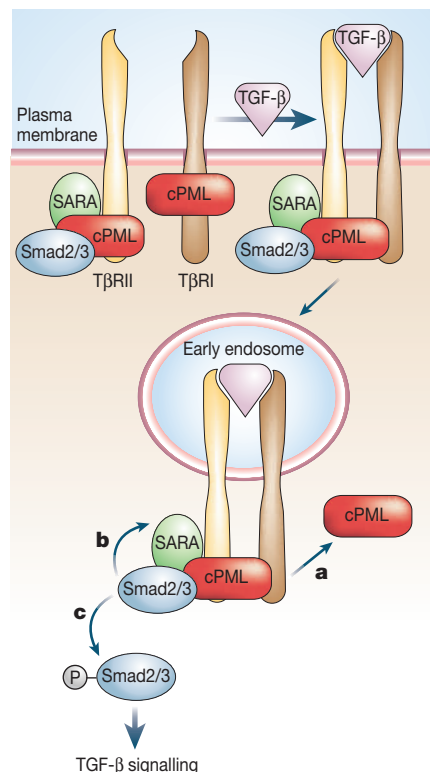
Cells communicate through signals that must be propagated from cell surface to nucleus. Tracking the signals generated by the transforming growth factor- $\beta$  protein reveals a surprising partner in this process.

During the development and maintenance of multicellular organisms, cells spend a lot of time communicating with each other to ensure that they are behaving in an appropriate and harmonious manner. This 'social' behaviour is mediated by the production of various signals outside the cell. These are interpreted in responding cells by networks of signal-transduction proteins that regulate the appropriate cellular response. One major class of signals emitted by cells in all animals is the transforming growth factor- $\beta$  (TGF- $\beta$ ) superfamily of growth and differentiation factors, which control normal development and biological processes, as well as cancer progression<sup>1</sup>. On page 205 of this issue, Lin and colleagues<sup>2</sup> report an unexpected participant in the network that interprets TGF- $\beta$  family signalling — cytoplasmic promyelocytic leukaemia protein (cPML).

PML is involved in a range of biological functions that include cell ageing, cell proliferation and programmed cell death<sup>3</sup>. As its name suggests, when these functions are disrupted, the result is acute promyelocytic leukaemia (APL) — so PML is classed as a tumour-suppressor protein. Most tumour-suppressing activities of PML are ascribed to its nuclear functions, but the function of the cytoplasmic form (cPML) was a mystery. Now it seems that it is a key accessory protein in the transduction of TGF- $\beta$  signals.

At the cell surface, TGF- $\beta$  interacts with two receptor complexes that function together to activate a downstream pathway of signalling proteins called Smads (Fig. 1). Smads move to the nucleus, where they can modulate cellular responses by regulating the expression of specific target genes<sup>4</sup>. One major feature of cell signalling is that receptors at the cell surface are transported into the cell, where they can either be recycled back to the cell surface, or be degraded. In the case of TGF- $\beta$  receptors, this internalization transfers some of the receptors into an organelle called the early endosome — a process that is crucial for TGF- $\beta$  signal transduction in some cells<sup>5</sup>. The importance of early endosomes in TGF- $\beta$  signalling is probably manifested by the presence of a protein called SARA (for 'Smad anchor of receptor activation') that can bind to both Smads and TGF- $\beta$  receptors<sup>6</sup>.

Lin *et al.*<sup>2</sup> report that cPML is crucial in orchestrating the dance that goes on during a signalling event, coordinating the receptors,



**Figure 1 Transforming growth factor- $\beta$  (TGF- $\beta$ ) signalling.** Lin *et al.*'s work<sup>2</sup> suggests a role for the cytoplasmic form of promyelocytic leukaemia protein (cPML). At the cell surface, cPML might interact with the two TGF- $\beta$  receptors (T $\beta$ RI and T $\beta$ RII) and act as a bridging factor between SARA and Smad2/3. Upon stimulation with TGF- $\beta$ , cPML promotes the transfer of the complex containing T $\beta$ RI, T $\beta$ RII, SARA and Smad2 into early endosomes. There, cPML might dissociate from the complex (a), allowing Smad2/3 to interact with SARA (b) and to be phosphorylated (c) by T $\beta$ RI. Phosphorylated Smad2/3 moves into the nucleus to propagate TGF- $\beta$  signalling.

SARA, Smads and the internalization process. They found that loss of cPML destroys TGF- $\beta$  signalling, and began to dissect the molecular pathways that underpin the process. One surprise to come out of the work is that cPML is required for the interaction of Smads with SARA. This is puzzling because SARA can bind directly to Smads<sup>7</sup>, so why is cPML necessary for the interaction *in vivo*? Another remarkable result from the paper provides a clue: cPML also binds to TGF- $\beta$  receptors and is required for the movement of both the TGF- $\beta$ -receptor complex and SARA into the early endosome. One possible implication is that

SARA can only bind to Smads inside the endosome, and that one role of cPML is to bring the two together. These findings suggest that interference with receptor localization and SARA-Smad interactions is the main mechanism by which loss of cPML abrogates cellular responses to TGF- $\beta$ .

The results raise lots of interesting questions, an obvious one being: how does cPML regulate the localization of receptors and SARA to the early endosome? Does it promote the transfer of the TGF- $\beta$  receptor into the endosome; does it regulate the time that the receptor complex stays in the endosome; or does it control the targeting of SARA to early endosomes? One hint at an answer comes from the observation that cPML has a RING-finger domain. In other cell-surface-located receptors, this domain mediates their trafficking by helping to attach a protein tag called ubiquitin to them<sup>8</sup>. So, one function of cPML might be to promote the internalization of the receptors.

The finding that cPML is required for TGF- $\beta$  signalling also converges nicely with observations that APL cell lines do not respond to TGF- $\beta$ <sup>2</sup>. APL is almost always linked with a rearrangement of the genes encoding PML and the retinoic-acid receptor- $\alpha$  (RAR $\alpha$ ). The outcome is a fusion involving one copy of the *PML* gene, which produces a PML-RAR $\alpha$  fusion protein; this has aberrant activity<sup>2</sup> and inhibits the activity of the normal PML in these tumours.

To wrap up their story, Lin *et al.* investigated TGF- $\beta$  resistance in APL cell lines that contain the PML-RAR $\alpha$  rearrangement. They confirmed that, in these cells, cPML-Smad complexes could not be detected and TGF- $\beta$  treatment does not promote Smad activation. By treating the cells with retinoic acid, which causes degradation of PML-RAR $\alpha$ , the authors reversed this effect, confirming that PML-RAR $\alpha$  does indeed confer TGF- $\beta$  resistance by interfering with normal cPML still present in the cell. Thus, the work also touches on how the loss of functional cPML might contribute to the 'antisocial' behaviour of cancers that display resistance to TGF- $\beta$ . Moreover, it reveals a link between anomalies in trafficking events in the TGF- $\beta$  system and alterations in the signalling properties of cancer cells.

Christine Le Roy and Jeffrey L. Wrana are in the Department of Medical Genetics and Microbiology, University of Toronto, and the Program in Molecular Biology and Cancer, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto M5G 1X5, Canada. e-mail: wrana@mshri.on.ca

1. Siegel, P. M. & Massagué, J. *Nature Rev. Cancer* **3**, 807–820 (2003).
2. Lin, H.-K., Bergmann, S. & Pandolfi, P. P. *Nature* **431**, 205–211 (2004).
3. Salomoni, P. & Pandolfi, P. P. *Cell* **108**, 165–170 (2002).
4. Attisano, L. & Wrana, J. L. *Science* **296**, 1046–1047 (2002).
5. Di Guglielmo, G. M. *et al. Nature Cell Biol.* **5**, 410–421 (2003).
6. Tsukazaki, T. *et al. Cell* **6**, 779–791 (1998).
7. Wu, G. *et al. Science* **287**, 92–97 (2000).
8. Marmor, M. D. & Yarden, Y. *Oncogene* **11**, 2057–2070 (2004).



## Obituary

## Edward B. Lewis (1918–2004)

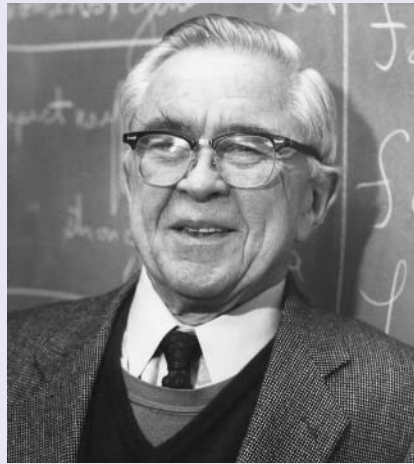
Ed Lewis, who died on 21 July at the age of 86, is remembered by all who knew him as a brilliant, eccentric and kindly scientist. He was a pioneer in exploring how genes design and build animals.

The 1995 Nobel Prize in Physiology or Medicine was a celebration of 85 years of study of the fruitfly *Drosophila*. The prize went to three biologists — Lewis, Christiane Nüsslein-Volhard and Eric Wieschaus — and we know now that the genes they discovered in *Drosophila* organize growth and development in all animals. Through decades of hard work, using endless patience and wonderful powers of observation, Lewis discovered and defined the first of these ‘designer’ genes. He called them the Bithorax complex; they give each body segment a distinct blueprint of instructions. If one segment’s blueprint is damaged, that segment produces components that really belong to a different segment. Thus, as Lewis impressively demonstrated, a mutant fly could be made with an extra pair of wings. It was later found that the genes of the Bithorax complex all have a DNA sequence called a homeobox, or Hox; this allowed the identification of equivalent genes in mammals and elsewhere.

Hox genes occur in clusters, and Lewis surprised everyone by finding that the arrangement of Hox genes in a cluster mirrors the order of the body parts, from head to tail, that those genes specify; this principle is common to all animals. But why? No one really knows — Lewis’s ideas still drive research. Lewis was also the first scientist to think in detail about where new genes come from. In 1951 he suggested that they arise by duplication of existing genes, one of which could then diverge from its parent. Molecular biologists confirmed this for Hox genes in the 1980s.

Lewis’s methods seem extremely simple: mate flies carrying different mutations, study the progeny, and then do more crosses. However, he was a master of both genetic logic and the details of fly anatomy, and, using both, he derived sophisticated insights into what genes are, how they are arranged and how they influence each other. He loved making models of gene regulation, creating ornate abstract constructs. These constructs with their special terminology were demanding, and by the time people had absorbed one, Lewis had moved on and built another.

To explain his ideas, he used animation. Being small, he could get under a table and move a magnet around, while, on top,



### Geneticist who pioneered studies of ‘designer’ genes in animal development

coloured pieces of metal sprung to life like whirligig beetles, imitating genes and their targets. We remember him at a conference in Crete, suffering from stage fright and equipped with a cardboard tube, bits of coloured plastic, three tennis balls and an obliging chairman, attempting unsuccessfully to explain an abstruse aspect of gene regulation. It was all good fun, but Lewis was serious about his work, for he was battling with life’s mysteries. The questions he defined still impinge on embryology, molecular biology, gene regulation, chromosome structure and evolution.

Lewis loved animals, harbouring snakes as a child. He and his wife Pam coddled their octopuses and desert tortoises. Lewis began work on *Drosophila* in high school, when a teacher encouraged him to study genetics. As an undergraduate at the University of Minnesota he worked with the geneticist C. P. Oliver, a student of Hermann Muller, who was himself a student of Thomas Hunt Morgan — the godfather of developmental genetics. As a graduate student at the California Institute of Technology (where, with a brief break during the Second World War, he would remain), Lewis’s mentor was Alfred Sturtevant, also a former student of Morgan’s. Thus Lewis was simultaneously the scientific grandson and great-grandson of Morgan.

Morgan began his career as an embryologist, but switched to *Drosophila* and genetics, leading the investigations that showed chromosomes to be the basis

of inheritance. Lewis completed the circle, using genetics to study development. Nevertheless, he did not start with this intention. Instead he aimed to explore the nature of the gene itself — “how it functions, how it mutates, how it evolves”.

His work also impinged on the study of cancer. In 1927, Muller reported that X-rays cause mutations in *Drosophila*. In 1955, Lewis discussed this work over lunch with some physicists, who had suggested that radiation damage to humans occurs only above a certain threshold dose. Lewis knew that Muller had found no such threshold in flies. Although it was not known then, Lewis also suspected that mutations can cause cancer. So he looked at data obtained from survivors of the Hiroshima and Nagasaki bombings, as well as from others exposed to radiation. He discovered a consistent dose–response relationship for leukaemia and other cancers, with no sign of a lower threshold. His papers questioned contemporary complacency about the safety of atomic-bomb testing and had considerable political and scientific impact. For his rigorously correct conclusions, he was attacked by the head of the United States Atomic Energy Commission.

For those who suspect that the present emphasis on publication is overdone, Lewis provides a superb role model. He published rarely and did not seem to care where. Some of his papers came out in such obscure journals that they were exchanged, like samizdat, as faded Xerox copies. His famously difficult 1978 paper in *Nature*, a compilation of decades of work on the Bithorax complex, finally brought him the wider attention he deserved.

A sweet, courteous and humble man, Ed worked in his lab to the end, probing possible connections between Hox genes and the cancer from which he was suffering. His credo, with which he ended his Nobel lecture, was: “Progress will still need to be driven by the logic of genetics and by further increases in abstraction,” following from his favourite quote from Bertrand Russell: “The power of using abstraction is the essence of intellect, and with every increase in abstraction the intellectual triumphs of science are enhanced.” **Matthew P. Scott and Peter A. Lawrence**

Matthew P. Scott is in the Departments of Developmental Biology, Genetics and Bioengineering, Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, California 94305-5439, USA.

e-mail: scott@pmgm2.stanford.edu

Peter A. Lawrence is in the Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK.

e-mail: pal@mrc-lmb.cam.ac.uk

## Chemical biology

### Synthetic stimulus to transcription

*J. Am. Chem. Soc.* **126**, 10504–10505 (2004)

A goal of the field known as chemical genetics is to control cell function using small synthetic molecules that interact with genes and proteins. As part of this endeavour, Aaron R. Minter *et al.* have identified a small molecule that can activate gene transcription.

Natural transcriptional activators are proteins that bind to specific DNA sequences and stimulate transcription of nearby genes. They typically possess a DNA-binding domain and an activation domain that mediates interactions between proteins (for example, by influencing the RNA polymerase that transcribes a gene).

Minter *et al.* looked for small molecular mimics of a 'minimal' peptide module, ATF14, found in a viral transcriptional activator called VP16. They synthesized several candidates containing the key functional groups of ATF14, attached them to a DNA-binding protein, and screened them for their ability to induce transcription of a 'test' gene. One of these molecules was nearly as active as ATF14 itself, despite being less than a fifth of its molecular weight. Such synthetic molecules are more resistant to proteolytic degradation than natural transcription factors — something that might partly account for the high activity of the best candidate here.

Philip Ball

## Atmospheric physics

### Raindrops on charge

*Geophys. Res. Lett.* **31**, doi:10.1029/2004GL020465 (2004)

Raindrops form when tiny cloud droplets coalesce after colliding with each other. Although the electrical charge on the droplets could regulate how quickly they come together, reliable information about charge size has been hard to come by.

Kenneth V. Beard *et al.* now report aircraft measurements of a droplet's charge from inside a cloud, data that have been difficult to collect given that an aircraft may itself charge up the droplets. To solve this problem, Beard *et al.* used a steady flow of nitrogen gas out of the sampling probe to keep any droplets that came into direct contact with the aircraft out of the sampling device.

For drops between about 10 and 26 micrometres across, they find that in different parts of a cloud the droplets can carry 80–90 positive or negative charges each. They point out that the electrical environment inside a cloud floating high above the Earth's surface is very different

from one drifting across a mountain top, where previous measurements revealed just a few charges per drop.

The measurements were carried out in low-lying stratocumulus clouds over Michigan in winter. The authors plan to extend their analyses to other seasons and cloud types, especially stratus and cumulus, in which ice and precipitation form more readily.

Mark Peplow

## Quantum mechanics

### Links and loops

*Phys. Rev. Lett.* **93**, 093001 (2004)

Topology enters quantum mechanics through the Aharonov–Bohm (AB) effect, in which an electromagnetic field affects a charged particle even if the particle moves along a trajectory for which the field is zero. For example, an electron executing a loop around a magnetic field can experience a phase shift of its quantum wavefunction, known as the Berry phase. John Kimball and Harry Frisch point out that such a situation can be realized within a topologically complex molecule or molecular assembly, and that the AB effect



could then lead to a measurable shift in the electronic energy levels of the molecules.

Consider, for example, a catenane, in which two ring-like molecules are interlinked. If one of the rings is electrically conducting and the other is magnetic (with its magnetization oriented along the chain), this is a molecular version of the AB experiment: an electron traversing the first link circumnavigates the magnetic flux of the second link. The energy shift then depends on the 'linking number' of the assembly, a purely topological quantity which specifies the number of times that the magnetic loop threads the conducting loop. A similar situation is seen in a single molecule if it is both conducting and magnetic and woven into a trefoil knot (pictured); the energy shift then depends on the 'writhe', a measure of the number of self-crossing points.

Philip Ball

## Neurobiology

### Sex and the brain

*Proc. Natl Acad. Sci. USA*  
doi:10.1073/pnas.0404644101 (2004)

It is not surprising that male and female brains are different, but the processes controlling certain neuroanatomical disparities between them are unclear. Early in development, the brains of both sexes are similar; by adulthood, however, differences in neuronal number can be seen in key brain regions. Nancy G. Forger and colleagues now implicate one of the main players in programmed cell death in controlling events in specific regions.

The team studied mice lacking the *Bax* gene, which is known to regulate the cell death of some developing neurons. Normally, female mice have more neurons than do males in an area of the brain called the anteroventral periventricular nucleus; males have more neurons in another area, the principal nucleus of the bed nucleus of the stria terminalis. There were no such sex differences in *Bax*-free mice.

It is thought that hormones, such as testosterone, help to control the survival of sex-specific neurons during development. But from the study by Forger *et al.* it seems that *Bax* can override hormonal signals, and may be the target gene regulated by testosterone.

Helen Pilcher

## Biomaterials

### Stamp it out

*Adv. Mater.* **16**, 1345–1348 (2004)

The ability to control the distribution of cells within three-dimensional scaffolds could facilitate the development of engineered tissues, and provide a tool for studying biological interactions *in vitro*. One way to achieve such spatial control would be to create compartments of different physical structure or chemical composition within a biological gel — as Min D. Tang and colleagues now describe.

These authors have modified the process of soft lithography to stamp a micrometre-scale pattern onto a substrate. The pattern is defined by one biopolymer hydrogel (an insoluble network of protein molecules that swells in water) deposited on another, type I collagen. This is a protein found in many tissue types and is the most abundant form of collagen in the human body.

The stamp is then removed under a solution of phosphate-buffered saline, to which a collagen precursor is added that triggers gelation when the temperature is subsequently raised. When the enzyme dispase is introduced, it eats away the patterned hydrogel, leaving an array of cavities. Tang *et al.* have shown that endothelial cells can be encapsulated within these cavities.

Rosamund Daw



# Parental care in an ornithischian dinosaur

A dramatic fossil may shed light on how modern archosaurs became devoted parents.

Crocodylians and birds show extensive parental care of their young<sup>1,2</sup>, but whether this behaviour evolved independently in these two groups of living archosaurs is unknown — in part because features of parenting among related fossil groups such as dinosaurs<sup>3</sup> are unclear. A dramatic specimen of the small ornithischian dinosaur *Psittacosaurus* sp. (Dalian Natural History Museum D2156) from Liaoning in China reveals a single adult clustered with 34 juveniles within an area of 0.5 square metres, providing strong evidence for post-hatching parental care in Dinosauria.

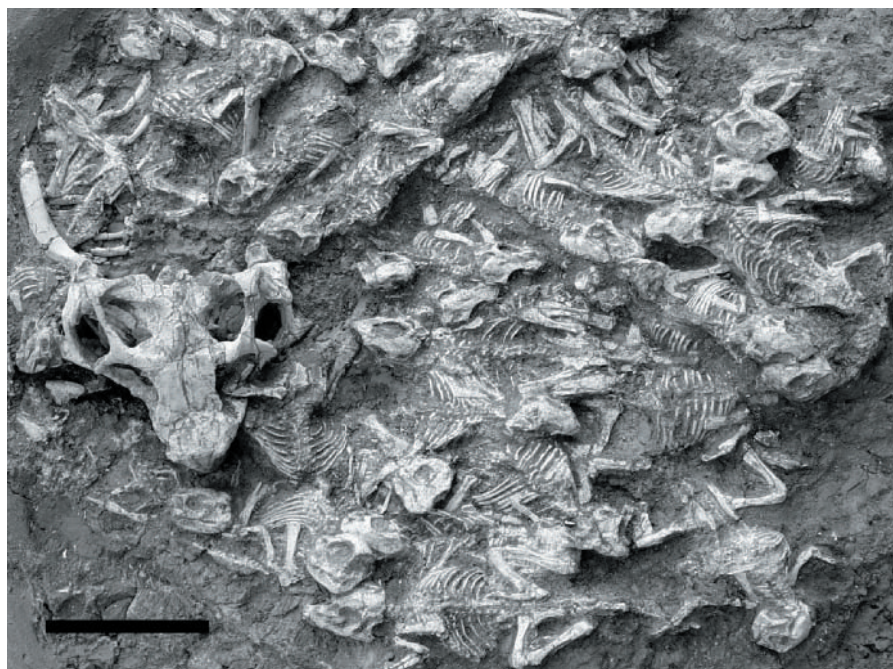
The Lower Cretaceous Yixian Formation, famous for its exceptionally well preserved feathered dinosaurs and birds, consists primarily of thinly layered lacustrine beds intercalated with volcanic deposits, as well as more massive units in the Lujiatun beds near its base<sup>4</sup>. The new specimen comes from a small outcrop of variegated mudstones, with root traces and gradational contacts, from the lowest portion of the formation near Shangyuan in western Liaoning.

The find consists of 34 articulated juvenile *Psittacosaurus* sp., closely associated with an adult (Fig. 1). The specimen is in red–grey mudstone, with clasts of volcanic rock fragments and minerals, feldspar and quartz. The unit is devoid of sedimentary structures apart from drab burrow mottling. Colour, composition, texture and sedimentary structures of this and adjacent units are consistent with the incipient development of almost neutral to slightly acidic soils on volcanically derived sediment.

There are no isolated bones or partial skeletons in the specimen, except those damaged by recent, pre-discovery erosion or excavation error. Juveniles show a consistent pattern of preservation: their skeletons have dorsal vertebrae, ribs and pelvis, with dorsal side up; the fore- and hindlimbs rest in lifelike poses, with humeri and femora splayed and the remaining portions folded below. The juveniles are all of similar size, with femora lengths ranging over 30–34 mm. Although postcrania commonly overlap one another, the skulls remain free of overlying skeletons.

The adult rests in a similarly articulated and upright pose. Visible elements include the skull, dorsal vertebrae, ribs and the dorsal portions of the scapula. The cervical series and posterior portion of the body are absent, probably lost to recent erosion. Portions of four juveniles overlies the rest of the skeleton but, like the juveniles, the adult head is slightly raised and uncovered.

In contrast to typical Yixian Formation skeletons such as *Sinosauropteryx*, which are



**Figure 1** Adult and juvenile *Psittacosaurus* in plan view. Erosion has truncated several skeletons, including the adult, on the left and upper sides of the specimen. Skeletons in the centre sit topographically lower than those at the perimeter, suggesting an original basin-like feature. One skeleton at lower right appears to be draped over an edge of this structure. The absence of internal sedimentary structures makes it impossible to discern whether the basin topography is the result of sedimentary, biological or post-depositional deformational processes. Juveniles not adjacent to the adult generally lie subparallel to one another, showing no preferred orientation of the cranial end. Sediment texture and composition do not vary across the site, eliminating the possibility of a composite specimen. Scale bar, 10 cm.

flattened between thin lacustrine strata<sup>4</sup>, these *Psittacosaurus* retain a three-dimensional form and occur in more massive units. This posture is frequently seen in other vertebrate specimens from the lower Lujiatun beds<sup>4</sup>. The lack of disarticulation, weathering, scavenging or other disturbance indicates minimal subaerial exposure. The consistently lifelike postures and skulls free of overlapping skeletons rule out significant transport and suggest that the *Psittacosaurus* were rapidly entombed while still alive. Death scenarios might include burial by volcanic debris<sup>5</sup>, entrapment in a collapsed underground burrow<sup>6</sup>, or flooding of a nest or other surface excavation. The absence of glass and the poorly sorted nature of these sediments favour a rapid clastic deposition, rather than burial by a volcanic ash fall. The uniformity of the entombing sediments, perhaps a result of soil development, prevents identification of any definite event.

The close association of the adult and juvenile skeletons is consistent with a biological relationship and post-hatching parental care. The temporal extent of this care depends upon the age difference between hatchlings and these juveniles. Hatchling size is unknown for *Psittacosaurus*. Nevertheless, juvenile bones from this specimen are fully

ossified and well formed. This, combined with bone development, slow growth rates<sup>7</sup>, plus the overall size and number of young, indicates that post-hatching growth was substantial and parental care extensive.

Parental care has been inferred for various dinosaur lineages from nestlings with altricial bone tissues, assemblages of shed teeth and tooth-marked bone, and mixed-age bone beds<sup>3,8</sup>, although these interpretations have been challenged on histological and taphonomic grounds<sup>9,10</sup>. The *Psittacosaurus* aggregation provides compelling evidence for post-hatching care among dinosaurs<sup>3</sup>.

Modern archosaurs are conscientious parents, assisting their young in hatching, protecting them from predators, feeding them and providing warmth and shelter<sup>12</sup>. But, given the disparity in ecology and physiology between crocodylians and birds, homology of their parental care is debatable<sup>11–13</sup>. Discovery of similar aggregations for other dinosaurs would strengthen the idea that post-hatching parental care is the ancestral condition in Dinosauria and therefore a homologous character among crocodylians and birds.

**Qingjin Meng\***, **Jinyuan Liu\***,  
**David J. Varricchio†**, **Timothy Huang‡**,  
**Chunling Gao\***

\*Dalian Natural History Museum, 40 Xicun Street,

Heishijiao, Shahekou, Dalian, China

†Department of Earth Sciences, Montana State University, Bozeman, Montana 59717, USA

e-mail: djv@montana.edu

‡PaleoWorld Research Foundation, Asian Operation B-1 22-2, Lane 256, RuiAn Street, Taipei, Taiwan 106, Republic of China

1. Magnusson, W. E., Vliet, K. A., Pooley, A. C. & Whitaker, R. in *Reproduction in Crocodiles and Alligators* (ed. Ross, C. A.) 118–135 (Facts on File, New York, 1989).
2. Gill, F. B. *Ornithology* (Freeman, New York, 1990).
3. Horner, J. R. *Annu. Rev. Earth Planet. Sci.* **28**, 19–45 (2000).
4. Zhou, Z., Barrett, P. M. & Hilton, J. *Nature* **421**, 807–814 (2003).
5. Voorhies, M. R. *Natl Geogr. Soc. Res. Rep.* **19**, 671–688 (1985).
6. Groenewald, G. H., Welman, J. & MacEachern, J. A. *Palaio* **16**, 148–160 (2001).
7. Erickson, G. M. & Tumanova, T. A. *Zool. J. Linn. Soc.* **130**, 551–566 (2000).
8. Bakker, R. T. & Bir, G. in *Feathered Dragons* (eds Currie, P. J., Koppelhus, E. B., Shugar, M. A. & Wright, J. L.) 301–342 (Indiana University, Bloomington, 2004).
9. Geist, N. R. & Jones, T. D. *Science* **272**, 712–714 (2003).
10. Carpenter, K. *Eggs, Nests, and Baby Dinosaurs* (Indiana University, Bloomington, 1999).
11. Tullberg, B. S., Ah-King, M. & Temrin, H. *Phil. Trans. R. Soc. Lond. B* **357**, 251–257 (2002).
12. Wesolowski, T. *Am. Nat.* **143**, 39–53 (1994).
13. Kavanau, J. L. *Lovebirds, Cockatiels, Budgerigars: Behavior and Evolution* (Science Software Systems, Los Angeles, 1987).

Competing financial interests: declared none.

## Bird song

# Superfast muscles control dove's trill

Bird songs frequently contain trilling sounds that demand extremely fast vocalization control<sup>1,2</sup>. Here we show that doves control their syrinx, a vocal organ that is unique to birds, by using superfast muscles. These muscles, which are similar to those that operate highly specialist acoustic organs such as the rattle of the rattlesnake, are among the fastest vertebrate muscles known and could be much more widespread than previously thought if they are the principal muscle type used to control bird songs.

The syrinx of ring doves (*Streptopelia risoria*) generates a relatively simple, highly stereotyped song — the familiar cooing sound. The coo contains a trill, whose elements are generated at repetition rates of up to 30 Hz (Fig. 1a). When doves coo, respiratory airflow excites membranes in the syrinx<sup>3</sup>, causing them to vibrate<sup>4,5</sup>. The vibrations depend on the tension in the membranes<sup>3,6</sup>, which is modified by activating two pairs of syringeal muscles<sup>5,7</sup>. These muscle pairs act as antagonists<sup>5</sup>: the tracheolateralis muscles cause the membranes to move out of the tracheal lumen (abduct), whereas the sternotrachealis muscles cause the membranes to slacken and to adduct.

We made simultaneous *in vivo* recordings of muscle activity and sound in cooing doves and found that the electromyographic (EMG) activity of the tracheolateralis muscles correlated significantly more strongly with voiced, rather than silent, periods: activating these muscles switches the sound

on, and not off as previously thought<sup>5</sup>, by positioning the membrane in the airflow. We also found that modulation of the tracheolateralis muscles' EMG correlated strongly with changes in sound frequency: the muscles change the tension in the membranes, which changes the frequency of the sound. (For methods, see supplementary information.)

A dove's trill cannot be achieved using typical vertebrate muscles, because they do not switch on and off fast enough to control the trill's brief sound elements ( $\geq 9$  ms). The syringeal muscles must also contract aerobically to power cooing sessions that can last for many minutes. These extreme requirements can be met only by aerobic superfast muscles<sup>8</sup>. This muscle type is the fastest known in vertebrates: its twitch half-time is less than 10 ms, which is one to two orders of magnitude faster than that of typical locomotory muscles<sup>8</sup>.

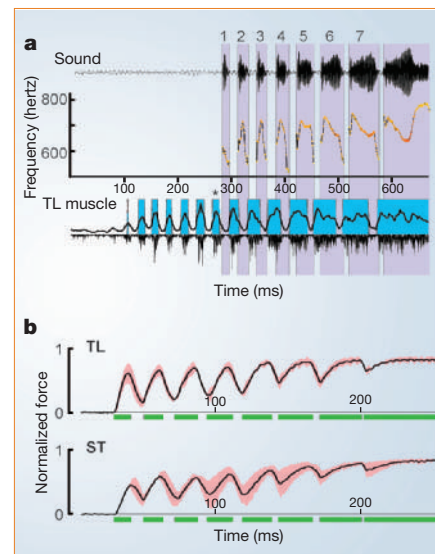
Our measurements show that dove syringeal muscles attain superfast kinetics. The twitch half-times of the tracheolateralis and sternotrachealis muscles are  $9.2 \pm 0.8$  ms and  $10.3 \pm 1.7$  ms, respectively, and the stimulus repetition rates necessary to obtain tetanic contraction are 200–275 Hz for tracheolateralis and 200–400 Hz for sternotrachealis muscles. But, like all superfast muscles, the syringeal muscles trade force for speed<sup>9</sup>: they exert low twitch stresses (tracheolateralis:  $8.0 \pm 4.8$  kilonewtons per  $m^2$ ; sternotrachealis:  $20.6 \pm 15.7$  kN  $m^{-2}$ ).

We also subjected the syringeal muscles to a playback signal based on the *in vivo* EMG pattern. During the simulated trill, both types of muscle modulate force (Fig. 1b). As the silent intervals in the playback signal shorten from 12 to 3.5 ms, neither is able to relax completely between trill elements.

Our results indicate that both muscles exert control directly at the syringeal membranes by altering the membrane position, which in turn alters membrane tension. The tracheolateralis muscles pull the syringeal membranes apart, thereby allowing them to vibrate in the airflow, which supports indirect evidence that the syrinx is closed between trills<sup>10</sup>. This control mode implies that gating is determined by the membrane position and that membrane tension determines pitch.

Our *in vitro* experiments show that both syringeal muscles have the superfast properties necessary to control individual sound elements during the dove's trill. Co-activation of the antagonistic muscles affords the bird very fast and accurate position control of the syringeal membranes, akin to saccadic eye and rapid arm movements<sup>11</sup>.

Birds modulate their songs extremely rapidly, with frequencies exceeding 100 Hz (ref. 2). Although the intrinsic nonlinear properties of the syrinx<sup>6</sup> add complexity to the level of motor control, only muscle control can explain the fast but gradual modulations that underlie the extraordinary intraspecific variability and flexibility of



**Figure 1** Superfast muscles control song sound in ring doves.

**a**, Sound oscillogram, spectrogram and electromyographic activity of tracheolateralis (TL) muscles (upward, integrated; downward, rectified) during the trill. TL activation (blue boxes) precedes the corresponding trill elements (purple boxes). Asterisk indicates TL burst associated with the first trill pulse (trill 1). **b**, Normalized traces of *in vitro* force modulation of TL and sternotrachealis (ST) muscles during a stimulation pattern (green bars) based on an averaged trill (mean (black trace)  $\pm$  s.d. (pink trace);  $n = 4$ ).

phonation<sup>1</sup>. The stereotyped coos of doves are considered to be simple vocalizations among birds, but even doves use superfast muscles to control their song. Given their added vocal complexity<sup>12</sup>, songbirds have probably evolved muscles that outperform the syringeal muscles of doves. Superfast muscle can no longer be considered a rare adaptation, found for example in the highly derived acoustic organs of the toadfish and rattlesnake<sup>8</sup>. We suspect that superfast vocal muscles are widespread among birds.

**Coen P. H. Elemans\***, **Igor L. Y. Spierts\***, **Ulrike K. Müller\***, **Johan L. van Leeuwen\***, **Franz Goller†**

\*Experimental Zoology Group, Wageningen University, 6709 PG Wageningen, The Netherlands e-mail: coen.elemans@wur.nl

†Department of Biology, University of Utah, Utah 84112, USA

1. Greenewald, C. H. *Bird Song: Acoustics and Physiology* (Smithsonian Institution, Washington DC, 1968).
2. Goller, F. & Suthers, R. A. *J. Neurophysiol.* **75**, 867–876 (1996).
3. Fletcher, N. H. *J. Theor. Biol.* **135**, 455–481 (1988).
4. Larsen, O. N. & Goller, F. *Proc. R. Soc. Lond. B* **266**, 1609–1615 (1999).
5. Goller, F. & Larsen, O. N. *J. Exp. Biol.* **200**, 2165–2176 (1997).
6. Fee, M. S., Shraiman, B., Pesaran, B. & Mitra, P. P. *Nature* **395**, 67–71 (1998).
7. Gardner, T., Cecchi, G., Magnasco, M., Laje, R. & Mindlin, G. B. *Phys. Rev. Lett.* **87**, 208101 (2001).
8. Young, I. S. & Rome, L. C. *Proc. R. Soc. Lond. B* **268**, 1965–1970 (2001).
9. Rome, L. C., Syme, D. A., Hollingworth, S., Lindstedt, S. L. & Baylor, S. M. *Proc. Natl Acad. Sci. USA* **96**, 5826–5831 (1999).
10. Beckers, G. B. L., Suthers, R. A. & ten Cate, C. *J. Exp. Biol.* **206**, 1833–1843 (2003).
11. Harris, C. M. & Wolpert, D. M. *Nature* **384**, 780–784 (1998).
12. Goller, F. *Nature* **395**, 11–12 (1998).

**Supplementary information** accompanies this communication on Nature's website.

Competing financial interests: declared none.



# High-resolution record of Northern Hemisphere climate extending into the last interglacial period

North Greenland Ice Core Project members\*

\*A full list of authors appears at the end of this paper

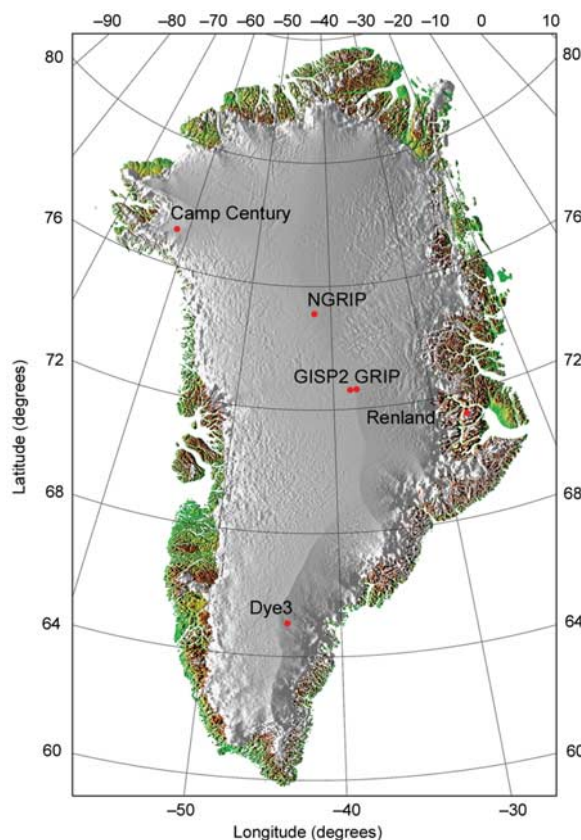
**Two deep ice cores from central Greenland, drilled in the 1990s, have played a key role in climate reconstructions of the Northern Hemisphere, but the oldest sections of the cores were disturbed in chronology owing to ice folding near the bedrock. Here we present an undisturbed climate record from a North Greenland ice core, which extends back to 123,000 years before the present, within the last interglacial period. The oxygen isotopes in the ice imply that climate was stable during the last interglacial period, with temperatures 5 °C warmer than today. We find unexpectedly large temperature differences between our new record from northern Greenland and the undisturbed sections of the cores from central Greenland, suggesting that the extent of ice in the Northern Hemisphere modulated the latitudinal temperature gradients in Greenland. This record shows a slow decline in temperatures that marked the initiation of the last glacial period. Our record reveals a hitherto unrecognized warm period initiated by an abrupt climate warming about 115,000 years ago, before glacial conditions were fully developed. This event does not appear to have an immediate Antarctic counterpart, suggesting that the climate see-saw between the hemispheres (which dominated the last glacial period) was not operating at this time.**

The two deep ice cores drilled at the beginning of the 1990s in central Greenland (GRIP<sup>1–3</sup> and GISP2<sup>4,5</sup>, respectively 3,027 m and 3,053 m long) have played a key role in documenting rapid climate changes during the last glacial period. However, it quickly became clear that the bottom 10% of at least one (and most probably both) of these ice cores<sup>4,6–9</sup> was disturbed owing to ice folding close to the bedrock. The Central Greenland ice core records are fully reliable climate archives back to 105,000 years before present (105 kyr BP), but the disturbances mean that no reliable Northern Hemisphere ice core record of the previous interglacial (the Eemian climatic period) was known to exist in the Northern Hemisphere.

This situation motivated the search for a new drilling site where undisturbed ice from the last interglacial period<sup>10</sup>, and even from the previous glacial period, would be accessible<sup>11</sup>. The North Greenland Ice Core Project (NGRIP) site, located at 75.10 °N and 42.32 °W with an elevation of 2,917 m and an ice thickness of 3,085 m (Fig. 1), was selected on the basis of three criteria that, when satisfied together, should produce dateable ice older than that found in central Greenland: a position on a ridge to reduce deformation by ice flow, flat bedrock, and a lower precipitation rate. The present accumulation rate is 0.19 m ice equivalent yr<sup>–1</sup>, the annual mean temperature is –31.5 °C, and the ice near the base originates 50 km upstream of the ice ridge in the direction of Summit<sup>12</sup>. The NGRIP drilling started in 1996, and bedrock was reached in July 2003.

## Dating of the NGRIP climate record

The climate record of the oxygen isotopic composition of the ice ( $\delta^{18}\text{O}$ ) from the NGRIP ice core is shown in Fig. 2 (and is available as Supplementary Information). In cold glaciers where the basal ice temperature is below freezing, the annual ice layers typically thin towards zero thickness close to bedrock, and flow induced disturbances can limit the usefulness of the deepest part of ice cores<sup>13</sup>. In contrast, at NGRIP high rates of basal ice melting, estimated to be 7 mm yr<sup>–1</sup> (refs 12, 14), remove the bottom layers, greatly restricting the thinning of the layers and the possibility of ice disturbances. Whereas the present-day accumulation is 15% lower at NGRIP than at GRIP, NGRIP annual layer thicknesses at 105 kyr BP (depth 2,900 m) are of the order of 1.1 cm, twice that of GRIP ice of this age.



**Figure 1** Map of Greenland, showing the locations of the deep ice core drilling sites. The sites GRIP (72.5 °N, 37.3 °W), GISP2 (72.5 °N, 38.3 °W), NGRIP (75.1 °N, 42.3 °W), Camp Century (77.2 °N, 61.1 °W), Dye3 (65.2 °N, 43.8 °W) and Renland (71.3 °N, 26.7 °W) are marked. The Greenland map was provided by S. Ekholm, Danish Cadastre.

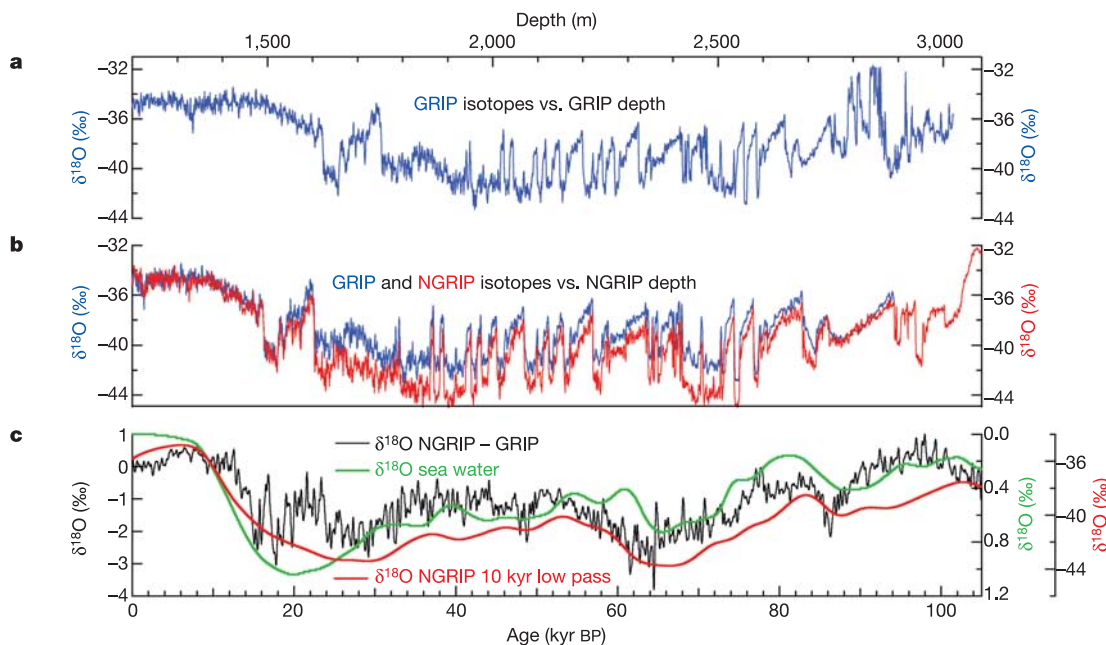
The NGRIP isotopic record covers the Holocene, the entire last glacial period, and part of the Eemian period. The 24 abrupt and climatic warm Dansgaard–Oeschger (DO) events, or Greenland interstadials (GIS), initially numbered in the GRIP record<sup>1,2</sup> are very clearly identified (Fig. 2a, b), as are the climatic cold Greenland stadials (GS) that follow the DO events. The NGRIP core has been cross-dated to the GRIP core ss09sea chronology<sup>15</sup> down to 105 kyr BP using the high-resolution ice isotope profiles and volcanic events found in the ECM and DEP records<sup>5,16</sup>. Older ice is cross-dated to the Antarctic Vostok ice core records by using concentrations of methane and  $\delta^{18}\text{O}$  of the entrapped air<sup>9,17–22</sup>. To determine if deep ice folding is a problem at NGRIP, we concentrate on the period corresponding to the marine isotope stage (MIS) 5d/5c transition dated around 105 kyr BP at Vostok (GT4 timescale). From methane and  $\delta^{15}\text{N}$  air measurements, we confirm that this transition is the counterpart of the Northern Hemisphere stadial 25<sup>18,21,23</sup> that ends with the abrupt onset of DO 24 at the NGRIP depth of 2,940 m (Fig. 3). At this depth, methane concentrations in air exhibit a rapid increase from 450 to 650 p.p.b.v., a shift which is also observed in the Vostok data<sup>24</sup> (Fig. 3), and the  $\delta^{15}\text{N}$  air signal, measured with a resolution of better than 100 yr, shows a rapid increase typical of DO events, resulting from thermal and gravitational fractionation processes. The increase in  $\delta^{15}\text{N}$  and in methane concentration over the warming of DO 24 are both located 7 m deeper in the ice core than the corresponding  $\delta^{18}\text{O}$  transition<sup>25–27</sup> (Fig. 3). This reflects the typical depth shift, or gas-age/ice-age difference, expected with normal firnification processes and later thinning through ice flow<sup>28</sup>. This supports our contention that the bottom ice is undisturbed by folding or ice mixing. We note that similar investigations on the GRIP core have confirmed that this record is indeed disturbed at the time of the 5d/5c transition<sup>7,18</sup>, as in that core the isotope and gas transitions are located at the same depth.

Below DO 19 the NGRIP record is compared to the planktonic oxygen isotope record from marine core MD95-2045 drilled on the

Iberian margin<sup>29</sup> (Fig. 4). On the basis of strong similarities between these two records and ice modelling as well as  $\delta^{18}\text{O}$  air measurements on the deepest parts of the core compared with Vostok, the basal part of the NGRIP record is dated to 123 kyr BP. Owing to the basal melting, the annual layer thickness of the ice from 2,700 to 3,085 m (90 to 123 kyr BP) thins much less than in the case of no melting, further making dating straightforward. At these depths, the depth scale is almost linearly proportional to time. Thus, we feel confident in interpreting the ice isotopic record at NGRIP as the first Northern Hemisphere ice core record of a highly detailed, undisturbed climate record of the late Eemian and the inception of the last glacial period.

### Climate record of the late Eemian period

We first examine the implications arising from the relatively high (warm) and stable Eemian ice isotopic values found in the bottom 85 m of the ice core. As noted above, the annual layers are unusually thick, 1.0 to 1.6 cm, through this period of glacial interception and the latter part of the Eemian period, allowing a very detailed look at this key climatic period. The maximum isotopic value of  $-32\text{‰}$  found for the Eemian in the NGRIP core corresponds to the highest values found in the GRIP and GISP2 ice cores. Although these other cores have disturbed chronologies for ice older than 105 kyr BP, they do contain Eemian age ice<sup>15,18</sup>, and the maximum isotopic values can be assumed to represent the warmest Eemian climate<sup>30</sup>. Because both the present interglacial isotopic values ( $-35\text{‰}$ ) and the Eemian values are similar in the GRIP, GISP2, and NGRIP ice, we infer that the ice from the bottom of the NGRIP core has sampled the warmest part of Eemian climate. This maximum isotopic value is  $3\text{‰}$  higher than the present value, and if attributed solely to temperature, implies at least a 5 K warmer temperature in the Eemian than at present<sup>30–33</sup>. It is notable that the  $3\text{‰}$  isotopic value difference between the present and the Eemian period seen at NGRIP, GRIP and GISP2 is also found in northern Greenland ice



**Figure 2** The NGRIP stable oxygen isotopic record compared to the GRIP record. **a**, The GRIP oxygen isotopic profile (blue) with respect to depth at GRIP. Isotopic values ( $\delta^{18}\text{O}$ ) are expressed in ‰ with respect to Vienna Standard Mean Ocean Water (V-SMOW). The measurements have been performed on 55 cm samples with an accuracy of  $\pm 0.1\text{‰}$ . **b**, The NGRIP oxygen isotopic profile (red) with respect to depth at NGRIP. For comparison,

the GRIP record (blue) has been plotted on the NGRIP depth scale using the rapid transitions as tie points. **c**, The difference between the NGRIP and GRIP oxygen isotopic profiles plotted above on the GRIP2001/ss09sea timescale<sup>15</sup> in 50 yr resolution (black). The record is compared to a record representing sea level changes<sup>39</sup> (green) and a 10-kyr smoothed oxygen isotope profile from NGRIP (red).



cores nearer the coast, such as Camp Century (77.2 °N, 61.1 °W) in the west<sup>15</sup>, and Renland (71.3 °N, 26.7 °W) in the east<sup>15</sup>. We conclude that the relative elevation differences during the Eemian in northern Greenland are thus not large, and further, as the Renland ice cap only is 325 m thick and cannot change elevation by more than 100 m, the absolute elevation changes between the Eemian and the present can only be of the order of 100 m. In contrast, the Dye3 ice core in south Greenland (65.2 °N, 43.8 °W) has an isotope difference of 5‰ (ref. 15), suggesting as much as 500 m lower elevation there. The Eemian isotopic values reported here paint a picture of an Eemian ice sheet with northern and central ice thicknesses similar to the present, while the south Greenland ice thickness is substantially reduced. This provides a valuable constraint for both future glaciological models of the Greenland Eemian ice sheet as well as models of sea level changes<sup>30,34–36</sup>.

### Climate record of the glacial inception

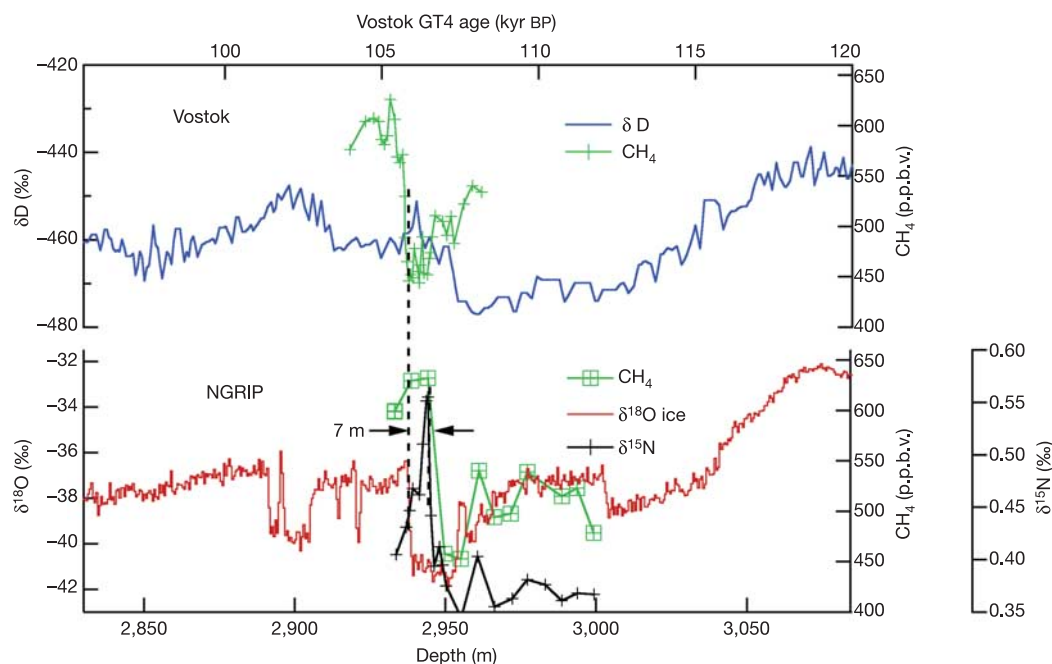
This high resolution NGRIP record reveals a slow decline in temperatures from the warm Eemian isotopic values to cooler, intermediate values over 7,000 yr from 122 to 115 kyr BP. The end of the last interglacial thus does not appear to have started with an abrupt climate change, but with a long and gradual deterioration of climate. Before full glacial values are reached, however, the record does reveal an abrupt cooling, with a first  $\delta^{18}\text{O}$  decrease at about 119 kyr BP, followed by relatively stable depleted  $\delta^{18}\text{O}$  levels, which we name here the Greenland stadial 26. The stadial is followed by an abrupt increase at ~115 kyr BP, the onset of DO 25<sup>37</sup> (Fig. 4). NGRIP is the first ice core climate record to so clearly resolve these rapid and large fluctuations in climate right at the beginning of the full glacial period. It is remarkable how well the features of the record compare with the marine planktonic isotope record from the margin of the Iberian coast, a proxy for the sea surface temperatures here. The features are thus believed to be large-scale features typical of the North Atlantic region<sup>38</sup>. It is significant that DO 25, while weak

(with an amplitude 25% of the following DO events), was similar in character to the following DO events, although it occurred at the time when the ice caps were first building up. Thus it seems difficult to call on melting ice or other large freshwater input to the North Atlantic to trigger this event, although clearly we need more information from this and future ice cores to fully understand this first abrupt climate change of the last glacial.

### Regional climate differences in Greenland

We now focus on a detailed comparison of the NGRIP  $\delta^{18}\text{O}$  ice profile with the GRIP ice isotopic record over their common part. Despite being only 325 km apart, these records have significant differences that illustrate the importance of regional variations in Greenland climate, even on quite long timescales. Figure 2b shows the NGRIP ice isotope profile. The GRIP record shown in Fig. 2a is plotted on the NGRIP depth scale using the DO events as references, so the two records can be compared. At first glance, the two records are very similar as expected, given the relative proximity of the cores. But closer inspection shows substantial differences between the records. Whereas NGRIP and GRIP have very similar  $\delta^{18}\text{O}$  levels during the Holocene, glacial isotopic levels in the NGRIP record are systematically depleted by 1‰ to 2‰. The difference between these isotopic profiles (Fig. 2c) reaches maxima at about 15–20 kyr BP, 25–30 kyr BP and 60–70 kyr BP. The magnitude of the difference appears to be related to the Northern Hemisphere climate curve, as represented by a smoothed version of the NGRIP record, such that colder conditions have larger differences (Fig. 2c). The difference curve also compares relatively well to the global sea level curve<sup>39</sup>, implying that the extent of the glacial continental ice sheets may help to explain the difference.

The difference curve only weakly traces the DO events, suggesting that the differences are not very well connected to processes operating on millennial timescales. A preliminary reconstruction of past temperatures based on the measured borehole temperatures



**Figure 3** Comparison of ice core records from NGRIP and Vostok for NGRIP depths 2,830 to 3,085 m. The isotopic composition,  $\delta\text{D}$ , of the ice (blue) and of methane in the air (green) for Vostok are on the top, and the isotopic composition,  $\delta^{18}\text{O}$ , of the ice (red), methane (green) and  $\delta^{15}\text{N}$  (black) of the air for NGRIP are on the bottom. A 50-kyr resolution NGRIP record is available as Supplementary Information. The detailed Vostok methane profile combines published data and recent measurements performed to

examine the 5d/5c transition at Vostok<sup>24</sup>. The Vostok and NGRIP data are reported on their own scales, namely the GT4 timescale for Vostok (top axis) and the depth scale for NGRIP (bottom axis). These two independent scales have been simply shifted in order to match the sharp methane shift in Vostok with the sharp NGRIP warming at 2,940 m. Furthermore, matching of the two scales should result in the estimated mean 1.1 cm annual layer thickness for the NGRIP profile.

at NGRIP supports this finding. Temperatures reconstructed at NGRIP during the Last Glacial Maximum are several degrees colder than those at GRIP and GISP2. The observed isotopic differences are large, given the relatively small distance between the two sites, and our finding that the two sites are believed to have only undergone small relative elevation changes during the glacial period<sup>34,40</sup>. Whereas the isotopic records in the central parts of East Antarctica<sup>41</sup> are rather similar and thus do not reveal large and significant climatically driven differences, the Greenland sites, located just 325 km apart, reveal major differences. Now that we are beginning to have a spatial distribution of deep ice core records, this brings into play a new source of palaeoclimatic information for these deep ice cores, that is, changes in geographical gradients with time. Our best theory is to postulate that the air masses reaching the two sites during the glacial had different sources. In response to the extent of the Laurentide ice sheet, sea ice and the extensive North Atlantic ice shelves, NGRIP has become further from the ocean, and may have seen a higher fraction of air coming over the northern side of the Laurentide ice sheet, bringing with it colder and more isotopically depleted moisture than GRIP might have seen<sup>42,43</sup>. Taken as a whole, the findings here suggest that the atmospheric water cycle over Greenland is substantially different between modern and glacial worlds.

### Basal water under the ice

When drilling was completed at NGRIP, basal water flooded the deepest 45 m of the bore hole. Although we knew from temperature profiles taken in 2001–02 that the base of the ice sheet was at or very near the pressure melting point, liquid water was not seen in radar profiles done during site selection. The melt rate at the base at NGRIP is 7 mm ice yr<sup>-1</sup>, so the geothermal heat flow appears to be as high as 140 mW m<sup>-2</sup> (70 mW m<sup>-2</sup> from latent heat, and 70 mW m<sup>-2</sup> conducted through the ice based on the measured bore temperature). This high geothermal heat flow value is atypical for Precambrian shields<sup>44</sup> believed to cover most of Greenland. The recent indications of bacterial life in and under Antarctic ice<sup>45</sup> have revealed that the Earth possibly contains a previously unrecognized cold biosphere that would be actively involved in biogeochemical processes. Thus Greenland, like Antarctica, is now known to have

liquid water at its base in some locations, water that awaits further study for basal sediment composition and evidence of life in a truly extreme environment.

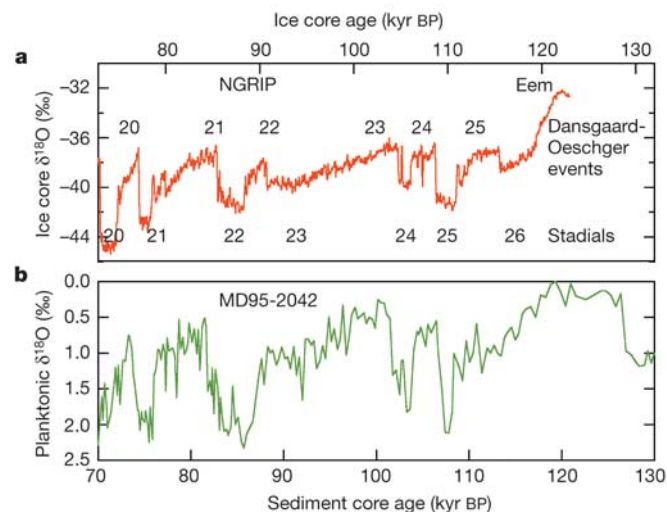
### Implications for future palaeoclimatic studies

The first measurements available on the NGRIP core already provide a wealth of new and promising environmental information. Most importantly, the NGRIP core contains the first continuous record of the late Eemian and the interception of the last glacial period to be recorded in a deep Greenland ice core. The palaeoclimatic signal for Greenland now reaches 123,000 yr back in time, and reveals a stable and warm late Eemian period. The end of the Eemian is a slow decline to glacial, cooler, intermediate conditions, but the onset of abrupt climate changes, the DO events that mark the last glacial period, precedes full glacial conditions. The bottom ice at NGRIP is essentially undisturbed and annual ice layers are quite thick, a situation caused by basal melting which in turn results from an unexpectedly high geothermal heat flow in North Greenland. The additional knowledge that the central and northern ice sheet during the Eemian period was at the same elevation as present constrains modelled ice volumes and sea level changes during the Eemian and glacial period. This interpretation is only consistent with modelling studies of the ice sheet during the Eemian that, although predicting an overall smaller ice sheet in accord with higher observed sea levels during this time<sup>34,35,46–48</sup>, allow for no large ice elevation change for the central Greenland ice. The next generation of models of the Greenland ice sheets should also include substantial melt under the northern part of the ice sheet as well as the northeast ice stream, important for the mass balance of the ice sheet<sup>49,50</sup>.

The deepest ice should allow a detailed study of the last glacial inception, including greenhouse gases and atmospheric dust loading, and in future comparisons with Antarctic records we should be able to investigate in detail the sequence of climatic events and forcing between north and south during this key climatic period. We find that the 5d/5c Vostok time period is the counterpart of the Northern Hemisphere stadial 25 that ends with the abrupt onset of DO 24 at the NGRIP depth 2,940 m. The north–south teleconnection observed here is similar in behaviour to all the following events (DO events 1–23), and behaves as predicted by the simple thermodynamic see-saw model<sup>23</sup>. In contrast, the weak stadial 26 followed by the abrupt onset of DO 25 is not opposed by an Antarctic reversal. This could be due to dating uncertainties between the two cores, but it could also be information on the timing of the onset of the teleconnection during the building of the ice caps and the cooling of the climate. When did the north–south climate see-saw begin? Is there information waiting to be found that can tell us how glacial periods begin, and whether we are in danger of entering one in the near future? New and detailed measurements from the EPICA Antarctica ice cores are expected to clarify this observation. And finally, is there life at the base of the Greenland ice sheet? These are some of the many questions that await further study of the new NGRIP ice core. □

Received 5 March; accepted 30 June 2004; doi:10.1038/nature02805.

1. Johnsen, S. J. *et al.* Irregular glacial interstadials recorded in a new Greenland ice core. *Nature* **359**, 311–313 (1992).
2. Dansgaard, W. *et al.* Evidence for general instability of past climate from a 250-kyr ice-core record. *Nature* **364**, 218–220 (1993).
3. Greenland Ice-Core Project (GRIP) Members. Climate instability during the last interglacial period recorded in the GRIP ice core. *Nature* **364**, 203–208 (1993).
4. Grootes, P. M., Stuiver, M., White, J. W. C., Johnsen, S. J. & Jouzel, J. Comparison of oxygen isotope records from the GISP2 and GRIP Greenland ice cores. *Nature* **366**, 552–554 (1993).
5. Taylor, K. C. *et al.* Electrical conductivity measurements from the GISP2 and GRIP Greenland ice cores. *Nature* **366**, 549–552 (1993).
6. Bender, M. *et al.* Climate correlations between Greenland and Antarctica during the last 100,000 years. *Nature* **372**, 663–666 (1994).
7. Fuchs, A. & Leuenberger, M.  $\delta^{18}\text{O}$  of atmospheric oxygen measured on the GRIP Ice Core document stratigraphic disturbances in the lowest 10% of the core. *Geophys. Res. Lett.* **23**, 1049–1052 (1996).
8. Johnsen, S. J. *et al.* The  $\delta^{18}\text{O}$  record along the Greenland Ice Core Project deep ice core and the



**Figure 4** The NGRIP isotopic profile from the Supplementary Information (a) compared with the planktonic isotopes in the Iberian margin sediment core MD95-2042<sup>29</sup> (b). The Greenland Dansgaard–Oeschger events (interstadials) are numbered along with the associated stadials. The two age scales are independent and seem to match within a few kyr.



- problem of possible Eemian climatic instability. *J. Geophys. Res.* **102**, 26397–26410 (1997).
9. Chappellaz, J., Brook, E., Blunier, T. & Malaizé, B. CH<sub>4</sub> and δ<sup>18</sup>O of O<sub>2</sub> records from Antarctic and Greenland ice: A clue for stratigraphic disturbance in the bottom part of the Greenland Ice Core Project and Greenland Ice Sheet Project 2 ice cores. *J. Geophys. Res.* **102**, 26547–26557 (1997).
  10. Dahl-Jensen, D. *et al.* The NorthGRIP deep drilling program. *Ann. Glaciol.* **35**, 1–4 (2002).
  11. Dahl-Jensen, D. *et al.* A search in north Greenland for a new ice-core drill site. *J. Glaciol.* **43**, 300–306 (1997).
  12. Dahl-Jensen, D., Gundestrup, N., Gorgineni, P. & Miller, H. Basal melt at NorthGRIP modeled from borehole, ice-core and radio-echo sounder observations. *Ann. Glaciol.* **37**, 207–212 (2003).
  13. Alley, R. B. *et al.* Comparison of deep ice cores. *Nature* **373**, 393–394 (1995).
  14. Grinsted, A. & Dahl-Jensen, D. A Monte Carlo-tuned model of the flow in the NorthGRIP area. *Ann. Glaciol.* **35**, 527–530 (2002).
  15. Johnsen, S. J. *et al.* Oxygen isotope and palaeotemperature records from six Greenland ice-core stations: Camp Century, Dye-3, GRIP, GISP2, Renland and NorthGRIP. *J. Quat. Sci.* **16**, 299–307 (2001).
  16. Clausen, H. B. *et al.* A comparison of the volcanic records over the past 4000 years from the Greenland Ice Core Project and Dye3 Greenland ice cores. *J. Geophys. Res.* **102**, 26707–26723 (1997).
  17. Blunier, T. *et al.* Asynchrony of Antarctic and Greenland climate change during the last glacial period. *Nature* **394**, 739–743 (1998).
  18. Landais, A. *et al.* A tentative reconstruction of the last interglacial and glacial inception in Greenland based on new gas measurements in the Greenland Ice Core Project (GRIP) ice core. *J. Geophys. Res.* **108**, doi:10.1029/2002JD0003147 (2003).
  19. Flückiger, J. *et al.* N<sub>2</sub>O and CH<sub>4</sub> variations during the last glacial epoch: Insight into global processes. *Glob. Biogeochem. Cycles* **18**, doi:10.1029/2003GB002122 (2004).
  20. Suwa, M., von Fischer, J. & Bender, M. Age reconstruction for the bottom part of the GISP2 ice core based on trapped methane and oxygen isotopes records. *Geophys. Res. Abstr.* **5**, 07811 (2003).
  21. Caillon, N. *et al.* Estimation of temperature change and of gas age-ice age difference, 108 kyr BP, at Vostok, Antarctica. *J. Geophys. Res.* **106**, 31893–31901 (2001).
  22. Petit, J. R. *et al.* Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature* **399**, 429–436 (1999).
  23. Stocker, T. F. & Johnsen, S. J. A minimum thermodynamic model for the bipolar seesaw. *Paleoceanography* **18**, doi:10.1029/2003PA000920 (2003).
  24. Caillon, N., Jouzel, J., Severinghaus, J. P., Chappellaz, J. & Blunier, T. A novel method to study the phase relationship between Antarctic and Greenland climate. *Geophys. Res. Lett.* **30**, doi:10.1029/2003GL017838 (2003).
  25. Severinghaus, J. P., Sowers, T., Brook, E. J., Alley, R. B. & Bender, M. L. Timing of abrupt climate change at the end of the Young Dryas interval from thermally fractionated gases in polar ice. *Nature* **391**, 141–146 (1998).
  26. Chappellaz, J. *et al.* Synchronous changes in atmospheric CH<sub>4</sub> and Greenland climate between 40 and 8 kyr BP. *Nature* **366**, 443–445 (1993).
  27. Brook, E. J., Sowers, T. & Orchardo, J. Rapid variations in atmospheric methane concentration during the past 110,000 years. *Science* **273**, 1087–1091 (1996).
  28. Schwander, J. *et al.* Age scale of the air in the Summit ice: Implication for glacial-interglacial temperature change. *J. Geophys. Res.* **102**, 19483–19493 (1997).
  29. Shackleton, N. J., Hall, M. A. & Vincent, E. Phase relationships between millennial-scale events 64,000–24,000 years ago. *Paleoceanography* **15**, 565–569 (2000).
  30. Cuffey, K. M. & Marshall, S. J. Substantial contribution to sea-level rise during the last interglacial from the Greenland ice sheet. *Nature* **404**, 591–594 (2000).
  31. Johnsen, S., Dahl-Jensen, D., Dansgaard, W. & Gundestrup, N. Greenland palaeotemperatures derived from GRIP bore hole temperature and ice core isotope profiles. *Tellus B* **47**, 624–629 (1995).
  32. Cuffey, K. M. *et al.* Large arctic temperature change at the Wisconsin-Holocene glacial transition. *Science* **270**, 455–458 (1995).
  33. Bennike, O. & Boecher, J. Land biotas of the last interglacial/glacial cycle, Jameson Land, East Greenland. *Boreas* **23**, 479–488 (1994).
  34. Marshall, S. J. & Cuffey, K. M. Peregrinations of the Greenland Ice Sheet divide in the last glacial cycle: Implications for central Greenland ice cores. *Earth Planet. Sci. Lett.* **179**, 73–90 (2000).
  35. Huybrechts, P. Sea-level changes at the LGM from ice-dynamic reconstructions of the Greenland and Antarctic ice sheets during the glacial cycles. *Quat. Sci. Rev.* **21**, 203–231 (2002).
  36. Gregory, J. M., Huybrechts, P. & Raper, S. C. B. Threatened loss of the Greenland ice-sheet. *Nature* **428**, 616 (2004).
  37. McManus, J. F., Oppo, D. W., Keigwin, L. D., Cullen, J. L. & Bond, G. C. Thermohaline circulation and prolonged interglacial warmth in the North Atlantic. *Quat. Res.* **58**, 17–21 (2002).
  38. Khodri, M. *et al.* Simulating the amplification of orbital forcing by ocean feedbacks in the last glacial. *Nature* **410**, 570–573 (2001).
  39. Waelbroeck, C. *et al.* Sea-level and deep water temperature changes derived from benthic foraminifera isotopic records. *Quat. Sci. Res.* **21**, 295–305 (2002).
  40. Reeh, N., Oerter, H. & Thomsen, H. H. Comparison between Greenland ice-margin and ice-core oxygen-18 records. *Ann. Glaciol.* **35**, 136–144 (2002).
  41. Watanabe, O. *et al.* Homogeneous climate variability across East Antarctica over the past three glacial cycles. *Nature* **422**, 509–512 (2003).
  42. Charles, C. D., Rind, D., Jouzel, J., Koster, R. D. & Fairbanks, R. G. Seasonal precipitation timing and ice core records. *Science* **269**, 247–248 (1995).
  43. Charles, C. D., Rind, D., Jouzel, J., Koster, R. D. & Fairbanks, R. G. Glacial-interglacial changes in moisture sources for Greenland: Influences on the ice core record of climate. *Science* **263**, 508–511 (1994).
  44. Dawes, P. R. in *Geology of Greenland* (eds Escher, A. & Watt, W. S.) 248–303 (The Geological Survey of Greenland, Denmark, 1976).
  45. Petit, J. R., Alekhina, I. & Bulat, S. A. in *Lessons for Exobiology* (ed. Gargaud, M.) (Springer, in the press).
  46. Letréguilly, A., Huybrechts, P. & Reeh, N. Steady-state characteristics of the Greenland ice sheet under different climates. *J. Glaciol.* **37**, 149–157 (1991).
  47. Letréguilly, A., Reeh, N. & Huybrechts, P. The Greenland ice sheet through the last glacial-interglacial cycle. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **90**, 385–394 (1991).
  48. Kukla, G., McManus, J. F., Rousseau, D.-D. & Chuine, I. How long and how stable was the Last Interglacial? *Quat. Sci. Rev.* **16**, 605–612 (1997).
  49. Fahnestock, M., Abdalati, W., Joughin, I., Brozena, J. & Coginini, P. High geothermal heat flow basal melt, and the origin of rapid ice flow in central Greenland. *Science* **294**, 2338–2342 (2001).
  50. Fahnestock, M. A. *et al.* Ice-stream-related patterns of ice flow in the interior of northeast Greenland. *J. Geophys. Res.* **106**, 34035–34045 (2001).

**Supplementary Information** accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** NGRIP is directed and organized by the Department of Geophysics at the Niels Bohr Institute for Astronomy, Physics and Geophysics, University of Copenhagen. It is supported by funding agencies in Denmark (SNF), Belgium (FNRS-CFB), France (IPEV and INSU/CNRS), Germany (AWI), Iceland (RannIs), Japan (MEXT), Sweden (SPRS), Switzerland (SNF) and the USA (NSF, Office of Polar Programs).

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to D.D.-J. (ddj@gfy.ku.dk) or S.J.J. (sigfus@gfy.ku.dk).

**K. K. Andersen<sup>1</sup>, N. Azuma<sup>2</sup>, J.-M. Barnola<sup>3</sup>, M. Bigler<sup>4</sup>, P. Biscaye<sup>5</sup>, N. Caillon<sup>6</sup>, J. Chappellaz<sup>3</sup>, H. B. Clausen<sup>1</sup>, D. Dahl-Jensen<sup>1</sup>, H. Fischer<sup>7</sup>, J. Flückiger<sup>4</sup>, D. Fritzsche<sup>7</sup>, Y. Fujii<sup>8</sup>, K. Goto-Azuma<sup>8</sup>, K. Grønvold<sup>9</sup>, N. S. Gundestrup<sup>1\*</sup>, M. Hansson<sup>10</sup>, C. Huber<sup>4</sup>, C. S. Hvidberg<sup>1</sup>, S. J. Johnsen<sup>1</sup>, U. Jonsell<sup>10</sup>, J. Jouzel<sup>6</sup>, S. Kipfstuhl<sup>7</sup>, A. Landais<sup>6</sup>, M. Leuenberger<sup>4</sup>, R. Lorrain<sup>11</sup>, V. Masson-Delmotte<sup>6</sup>, H. Miller<sup>7</sup>, H. Motoyama<sup>8</sup>, H. Narita<sup>12</sup>, T. Popp<sup>13</sup>, S. O. Rasmussen<sup>1</sup>, D. Raynaud<sup>3</sup>, R. Rothlisberger<sup>4</sup>, U. Ruth<sup>7</sup>, D. Samyn<sup>11</sup>, J. Schwander<sup>4</sup>, H. Shoji<sup>14</sup>, M.-L. Siggard-Andersen<sup>1</sup>, J. P. Steffensen<sup>1</sup>, T. Stocker<sup>4</sup>, A. E. Sveinbjörnsdóttir<sup>15</sup>, A. Svensson<sup>1</sup>, M. Takata<sup>2</sup>, J.-L. Tison<sup>11</sup>, Th. Thorsteinsson<sup>16</sup>, O. Watanabe<sup>8</sup>, F. Wilhelms<sup>7</sup> & J. W. C. White<sup>13</sup>**

**Affiliations for authors:** 1, Niels Bohr Institute for Astronomy, Physics and Geophysics, University of Copenhagen, Juliane Maries Vej 30, DK-2100 Copenhagen OE, Denmark; 2, Nagaoka University of Technology, 1603-1 Kamitomioka-machi, Nagaoka 940-2188, Japan; 3, Laboratoire de Glaciologie et Géophysique de l'Environnement (CNRS), BP 96, 38402 St Martin d'Hères Cedex, France; 4, Climate and Environmental Physics, Physics Institute, University of Bern, Sidlerstrasse 5, CH-3012, Switzerland; 5, Lamont-Doherty Earth Observatory of Columbia University, Rte 9W - PO Box 1000, Palisades, New York 10964-8000, USA; 6, Institute Pierre Simon Laplace/ Laboratoire des Sciences du Climat et de l'Environnement, UMR CEA-CNRS, CE Saclay, Ommes des Merisiers, 91191 Gif-sur-Yvette, France; 7, Alfred-Wegener-Institute for Polar and Marine Research (AWI), Postfach 120161, D-27515 Bremerhaven, Germany; 8, National Institute of Polar Research, Kaga 1-9-10, Itabashi-ku, Tokyo 173-8515 Japan; 9, Nordic Volcanological Institute, Grensávegur 50, 108 Reykjavík, Iceland; 10, Department of Physical Geography and Quaternary Geology, Stockholm University, S-106 91, Stockholm, Sweden; 11, Département des Sciences de la terre et de l'Environnement, Faculté des Sciences, CP 160/03, Université Libre de Bruxelles, 50 avenue FD Roosevelt, B1050 Brussels, Belgium; 12, Research Institute for Humanity and Nature, 335 Takashima-cho, Marutamachi-dori Kawaramachi nishi-iru, Kamigyo-ku, Kyoto 602-0878, Japan; 13, INSTAAR, Campus Box 450, University of Colorado, Boulder, Colorado 80309-0450, USA; 14, Kitami Institute of Technology, Koencho 165, Kitami, Hokkaido 090-8507 Japan; 15, Raunvísindastofnun Háskólans, Dunhagi 3, Iceland; 16, National Energy Authority, Grensávegur 9, IS-108 Reykjavík, Iceland

\*Deceased

# The ring of life provides evidence for a genome fusion origin of eukaryotes

Maria C. Rivera<sup>1,3,4</sup> & James A. Lake<sup>1,2,4</sup>

<sup>1</sup>Molecular Biology Institute, MCD Biology, <sup>2</sup>Human Genetics, <sup>3</sup>IGPP, and <sup>4</sup>Astrobiology Institute, University of California, Los Angeles 90095, USA

Genomes hold within them the record of the evolution of life on Earth. But genome fusions and horizontal gene transfer seem to have obscured sufficiently the gene sequence record such that it is difficult to reconstruct the phylogenetic tree of life. Here we determine the general outline of the tree using complete genome data from representative prokaryotes and eukaryotes and a new genome analysis method that makes it possible to reconstruct ancient genome fusions and phylogenetic trees. Our analyses indicate that the eukaryotic genome resulted from a fusion of two diverse prokaryotic genomes, and therefore at the deepest levels linking prokaryotes and eukaryotes, the tree of life is actually a ring of life. One fusion partner branches from deep within an ancient photosynthetic clade, and the other is related to the archaeal prokaryotes. The eubacterial organism is either a proteobacterium, or a member of a larger photosynthetic clade that includes the Cyanobacteria and the Proteobacteria.

The origin of the eukaryotic cell is enigmatic and complex. Early studies of nuclear-encoded enzymes, transfer RNAs, ribosome structures and ribosomal RNA catalogues implied deep, but unresolved, connections between prokaryotes and eukaryotes<sup>1–3</sup>. Subsequent analyses of ribosomal sequences suggested that eukaryotes were either the sister group to the Archaea<sup>4</sup>, or the sister group to the archaeal Eocyta<sup>5,6</sup> in the rooted tree of life<sup>7,8</sup>. But as additional eukaryotic genes were sequenced, their analyses frequently suggested just the opposite; that is, that many eukaryotic genes were more closely related to the Bacteria than to the Archaea<sup>9–12</sup>. A further complication arose from the finding that the function of each eukaryotic gene is strongly correlated with its ancestry. Informational genes (genes involved in transcription, translation and other related processes) are most closely related to archaeal genes, whereas operational genes (genes involved in cellular metabolic processes such as amino acid biosynthesis, cell envelope and lipid synthesis, and so on) are most closely related to eubacterial genes<sup>13</sup>. Recently, comprehensive analyses involving large numbers of genomes<sup>14</sup> have shown the strength of this correlation.

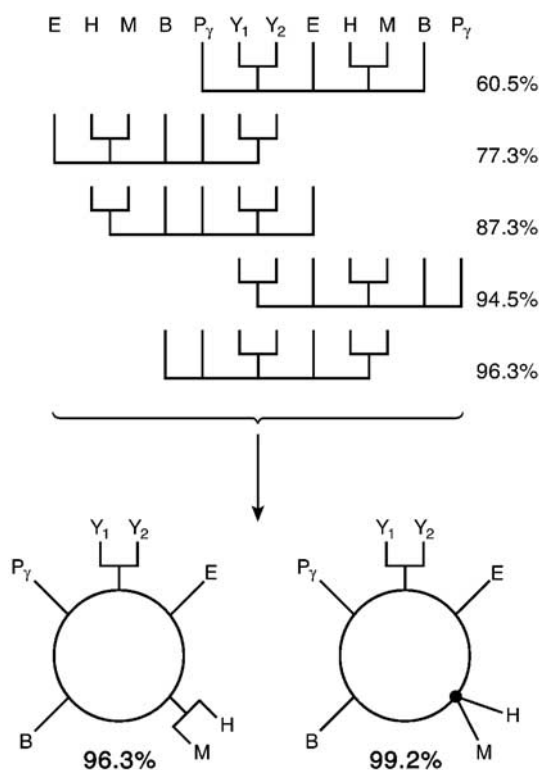
It has been difficult to reconcile these conflicting results with the origin of eukaryotes because of the complicating effects of genome fusions and horizontal (or lateral) gene transfer (HGT) on phylogenetic reconstructions. Genome fusions convert trees into rings, which cannot be analysed by conventional phylogenetic algorithms. Although multiple, independent events of HGT cannot make rings, the extensive horizontal transfer observed among prokaryotes<sup>13,15–19</sup> can obscure the identities of those prokaryotes that may have contributed genes to eukaryotes.

Recently a new algorithm, conditioned reconstruction, based on the two character states of gene presence and absence has been developed, which can reconstruct genome fusions<sup>20</sup>. Analyses of the presence or absence of genes have been used previously to reconstruct phylogenetic trees<sup>21–26</sup>, but these methods cannot detect genome fusions. Conditioned reconstructions, when used in conjunction with Markov-based quartet methods, can rigorously analyse the fusion of two genomes and ameliorate biases due to HGT. Here we apply this new method to investigate the evolution of eukaryotes, and present evidence that the eukaryotic genome arose through the fusion of two diverse prokaryotic genomes.

## The topology of the ring of life

Ten complete genomes from prokaryotic and eukaryotic organisms, comprising a representative sampling of the diversity of life, were analysed using the method of conditioned reconstruction to obtain

a better understanding of eukaryotic origins. (See Supplementary Data for the analyses of 24 further prokaryotic genomes and additional control experiments.) The five most probable trees from a set of three Bacteria, three Archaea and two eukaryotes are shown in Fig. 1 (the names of the taxa are listed in the legend). The

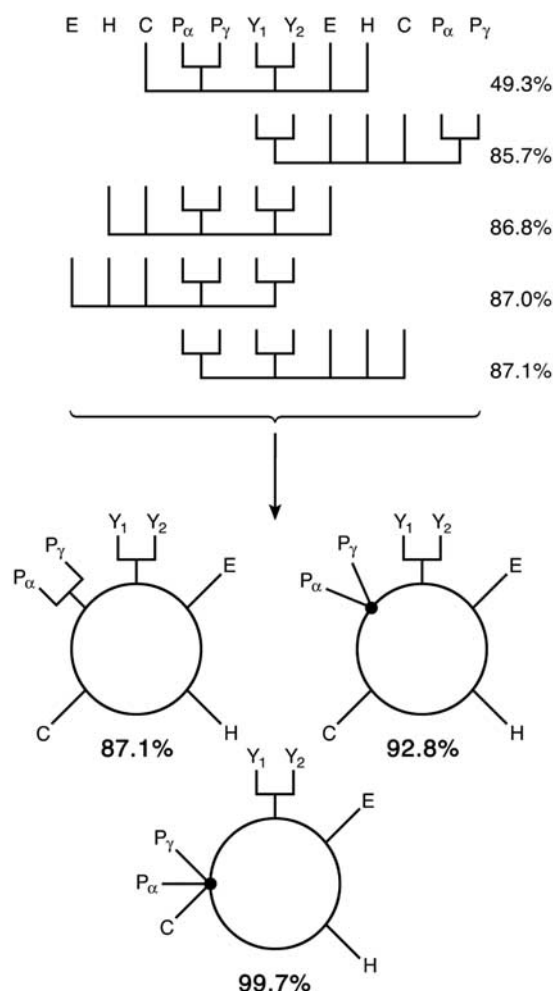


**Figure 1** Conditioned reconstructions provide evidence for the ring of life. The genomes are from two yeasts (Y<sub>1</sub>, *Schizosaccharomyces pombe* and Y<sub>2</sub>, *Saccharomyces cerevisiae*), a  $\gamma$ -proteobacterium (P<sub>γ</sub>, *Xylella fastidiosa*), a bacillus (B, *Staphylococcus aureus* MW2), a halobacterium (H, *Halobacterium* sp. NRC-1), a methanococcus (M, *Methanosarcina mazei* Goe1), an eocyte (E, *Sulfolobus tokodaii*) and an archaeoglobium (not shown; the conditioning genome *Archaeoglobus fulgidus* DSM4304). The five most probable unrooted trees are shown with leaves pointing upward to emphasize that each is part of a repeating pattern. Cumulative probabilities are shown at the right of each tree. Fully and partially resolved rings are at the lower left and right, respectively.



cumulative probabilities of these five trees are shown at the right of each tree. The upper tree is supported by 60.5% of the conditioned reconstruction bootstraps, the second tree by 16.8% of the conditioned reconstruction bootstraps, for a cumulative probability of 77.3%, and so on. It initially appears that the resolution of the tree is poor because the most probable tree is supported by a low bootstrap value, and the other trees are supported by even lower values. However, when the five most probable unrooted trees are aligned by shifting each to the left or the right, until their leaves match, they form a repeating pattern indicating that the five trees are simply permutations of an underlying cyclic pattern. This implies that they are derived from the single cycle graph (ring) shown at the bottom left of Fig. 1 (ref. 20). When that ring is cut at any one of the five central arcs and then unfolded, the resulting unrooted tree will correspond to one of the five most probable trees. In other words, the data are not tree-like they are ring-like.

A rigorous combinatorial analysis of the genomic fusion of two organisms valid for all possible initial pre-fusion trees has shown that the conditioned reconstruction algorithm recovers all permutations of the cycle graph, and only those permutations<sup>20</sup>. Hence, these results may be interpreted, in a manner analogous to the



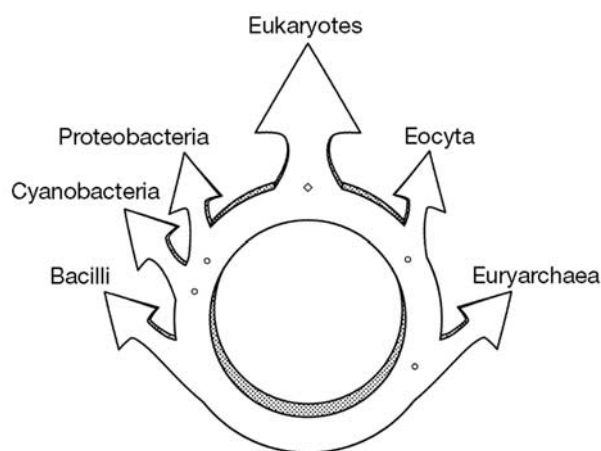
**Figure 2** Eubacterial relationships within the ring of life. The genomes are from two yeasts ( $Y_1$ , *S. pombe* and  $Y_2$ , *S. cerevisiae*), a  $\gamma$ -proteobacterium ( $P_\gamma$ , *X. fastidiosa* 9a5c), an  $\alpha$ -proteobacterium ( $P_\alpha$ , *Brucella melitensis* 16M), a cyanobacterium (C, *Synechocystis* sp. PCC6803), a halobacterium (H, *Halobacterium* sp. NRC-1), an eocyte (E, *S. tokodaii*) and a bacillus (not shown; the conditioning genome *S. aureus* MW2). The five unrooted trees consistent with the ring are shown with leaves pointing upward to emphasize that each is part of a repeating pattern. Cumulative probabilities are at the right of each tree. Fully and partially resolved rings are at the lower left, right and centre, respectively.

interpretation of restriction digests of a circular plasmid or the mapping of a circular chromosome, as implying a ring of life. The ring shown at the lower left of Fig. 1 is completely consistent with all five of the fully resolved trees shown at the top of the figure, hence we refer to it as a fully resolved ring. That ring explains 96.3% of the bootstrap replicates, and the partially resolved ring at the lower right of Fig. 1 explains 99.2% of the bootstrap replicates. The ring is supported for all possible conditioning genomes (see Supplementary Data). This provides strong evidence for the completely resolved ring on the left, and even stronger evidence for the less-resolved ring on the right. From these results we infer that the eukaryotic nuclear genome was formed from the fusion of the genomes of a relative of a proteobacterium ( $P_\gamma$ ) and a relative of an archaeal eocyte (E).

### Eubacterial roots of the ring of life

By analysing additional taxa, one can discover whether the eubacterial fusion partner branches even deeper than the  $\gamma$ -proteobacterium *Xylella*. The data set for this second analysis, described in the legend for Fig. 2, specifically included the  $\alpha$ -proteobacterium *Brucella* and the cyanobacterium *Synechocystis*. The results are shown in Fig. 2. The cumulative probability of the five trees, shown at the top of the figure, corresponds to 87.1% of the total bootstraps, indicating significant statistical support for the ring shown at the lower left of Fig. 2. The percentage increases to 92.8% when the  $\alpha$ -proteobacterium and the  $\gamma$ -proteobacterium are allowed to form an unresolved bifurcation, and rises to 99.7% when the  $\alpha$ -proteobacterium,  $\gamma$ -proteobacterium and cyanobacterium are allowed to form an unresolved trifurcation. The ring is also supported for all possible conditioning genomes (see Supplementary Data). The combined analyses in Figs 1 and 2 provide strong evidence for the ring in Fig. 3, and indicate that the immediate relative of the eubacterial fusion partner was probably a primitive proteobacterium, or possibly a primitive photosynthetic eubacterium.

The second analysis, illustrated in Fig. 2, did not detect an additional ring connecting the  $\alpha$ -Proteobacteria, the phylum from which mitochondria arose<sup>27</sup>, to the eukaryotes. There is increasing evidence that eukaryotes have received a deluge of DNA from organelles<sup>28</sup>. In one study, 630 nuclear-encoded genes in eukaryotes



**Figure 3** A schematic diagram of the ring of life. The eukaryotes plus the two eukaryotic root organisms (the operational and informational ancestors) comprise the eukaryotic realm (see Supplementary Discussion). Ancestors defining major groups in the prokaryotic realm are indicated by small circles on the ring. The Archaea<sup>49</sup>, shown on the bottom right, includes the Euryarchaea, the Eocyta and the informational eukaryotic ancestor. The Karyota<sup>5</sup>, shown on the upper right of the ring, includes the Eocyta and the informational eukaryotic ancestor. The upper left circle includes the Proteobacteria<sup>49</sup> and the operational eukaryotic ancestor. The most basal node on the left represents the photosynthetic prokaryotes and the operational eukaryotic ancestor.

were identified as having mitochondrial antecedents<sup>29</sup>. A substantial and continuous influx of mitochondrial DNA to the eukaryotic nucleus has been documented<sup>28–33</sup>, and the  $\alpha$ -proteobacterial relatives of mitochondria might have undergone HGT with other prokaryotic groups<sup>16–19</sup>. Hence the present data do not reveal whether the eubacterial fusion partner was distinct from the ancestor of mitochondria or identical with it. Further lineage sampling with conditioned reconstruction among  $\alpha$ -Proteobacteria and other prokaryotes might help to resolve this issue.

## Identifying the fusion organism

We have, so far, interpreted these results as strongly supporting the conclusion that two prokaryotes fused their genomes, thereby closing the ring of life and creating the first eukaryote. However the formal alternative exists that it was a prokaryote rather than a eukaryote that resulted from the genome fusion and closed the ring of life. Hence we explicitly tested the identity of the fusion organism as follows.

In tree analyses, each leaf contacts only one node of the tree, so that eliminating one taxon from the analysis will delete that leaf, but not otherwise change the tree. Similarly, in a conditioned reconstruction analysis, eliminating a non-fusion organism from the analysis will delete that leaf from the tree, or ring, without affecting the ring. However, eliminating a fusion organism, which necessarily contacts two nodes of a ring, will delete the leaf, open the ring and convert it into a tree<sup>20</sup>. When the taxa that comprise the ring in Fig. 2 were systematically removed, the ring opened only when both yeast genomes were simultaneously removed, indicating that the eukaryotic genome had inherited genes from its prokaryotic fusion partners. When the eukaryotes were removed, the fully resolved tree (Fig. 2, lower left) minus eukaryotes was supported by 88.6% of the bootstraps. The poorly resolved tree (Fig. 2, bottom) minus eukaryotes was supported by 99.96% of the bootstraps, thereby demonstrating that the yeast lineage is the fusion product of prokaryotes, as illustrated in Fig. 3. Furthermore, the conditioned reconstruction analysis of an additional 24 prokaryotic genomes in the absence of eukaryotic genomes reconstructed only trees, and not rings (Supplementary Fig. S1). This adds further support to the conclusion that eukaryotes are indeed the products of genome fusions.

## Discussion

Various theories have been proposed for the origin of the nuclear genes of eukaryotes. These include autogenous theories, chimaeric theories and genome fusion theories. The results derived in this paper argue against autogenous theories; that is, tree of life theories in which eukaryotes are proposed to have evolved clonally from a single, possibly very ancient, prokaryote. In this paper chimaeric theories refer to the acquisition of genes by eukaryotes from multiple sources through unspecified mechanisms. The data presented here argue against them, except of course for chimaeric theories that specifically propose genome fusions.

Genome fusion theories, in which eukaryotic nuclear genes are obtained through the fusion of two diverse genomes, are strongly supported by the analyses presented here. By default, an endosymbiosis between two prokaryotes is probably the mechanism responsible for the genome fusion described here, although the fusion signal may have been augmented by gene contributions from eukaryotic organelles. Symbiotic relationships are fairly common among organisms living together, and in rare cases this leads to endosymbiosis, the intracellular capture of former symbionts<sup>34</sup>. We and others have previously proposed endosymbiotic theories for the origin of eukaryotes<sup>26,35–39</sup>. Given a genome fusion, and in the absence of other mechanisms that could produce fusions, we conclude that an endosymbiosis was the probable cause.

The ring of life is consistent with, and confirms and extends, a number of previously reported results. It implies that prokaryotes

pre-date eukaryotes, as two pre-existing prokaryotes contributed their genomes to create the first eukaryotic genome. This probably places the root of the ring below the eubacterial–eukaryotic last common ancestor and the eocytic–eukaryotic last common ancestor, as shown in Fig. 3. This partial rooting of the ring of life is consistent with the eukaryotic rooting implied by the EF-1 $\alpha$  insert that is present in all known eukaryotic and eocytic EF-1 $\alpha$  sequences and lacking in all paralogous EF-G sequences<sup>12,40</sup>. Given the limited sampling here, the prokaryotic connections in the ring are broadly consistent with the results from insert distributions and concatenated gene trees<sup>12,41–43</sup>.

The ring of life explains a number of confusing observations that motivated this study. Because the eukaryotic genome resulted from a fusion it is expected that in some gene trees eukaryotes will be related to Bacteria, whereas in other gene trees eukaryotes will be related to Archaea, in accord with the results of others<sup>9–12</sup>. Our observations and those of others<sup>13,14</sup> suggesting that the informational genes of eukaryotes are primarily derived from Archaea and the operational genes are primarily derived from Bacteria are also consistent with the ring of life. Those observations suggest that the operational genes have come from the eubacterial fusion partner and the informational genes from the archaeal fusion partner. The ring of life does not explain why this happened, but it does provide a broad phylogenetic framework for testing theories for the origin and evolution of the eukaryotic genome. □

## Methods

Conditioned reconstructions<sup>20</sup> use Markov-based methods to determine phylogenetic trees and graphs based on the two character states 'gene presence' and 'gene absence' defined by gene orthologue sets. Gene orthologue sets used in these reconstructions were calculated in a three-step process to ensure that each gene orthologue set was (1) a globally optimal, strongly connected digraph, with orthologues explicitly distinguished from (2) paralogues and (3) recent duplications.

Globally optimal orthologue sets were constructed using gapped BLAST analyses (BLASTP, v.2) using the BLOSUM62 matrix and default parameters<sup>44</sup>. Because the top BLAST score is often not the most closely related orthologue<sup>41</sup>, we used the best three scores to construct preliminary orthologue sets (for details see ref. 20). Each gene orthologue set was defined by using a specific gene in the conditioning genome for reference. For example, if *Escherichia coli* was used as the conditioning genome, and the EF-Tu gene was used for reference, each non-*E. coli* genome would be searched for the top three BLAST hits for *E. coli* EF-Tu. From this collection, 2,187 unique sets of orthologues ( $3^{n-1}$  where  $n$  equals eight genomes total) were constructed. The sum of all possible pairwise BLAST scores between orthologues was calculated for each set. The set with the highest sum was chosen as a tentative orthologue set. In the second step of the procedure, gene paralogues were detected through computations based on the properties of digraphs. We have observed that paralogues are frequently unilaterally or even weakly connected to other nodes in a BLAST-based digraph. Hence each tentative orthologue set was systematically analysed to determine all non-empty, strongly connected subsets of the tentative orthologue set. These subsets of putative orthologous genes were ordered according to the sum of their reciprocal BLAST scores, from largest to smallest. The final step identified recently duplicated genes based on the observation that they are separated from each other by less time than they are from other members of an orthologue set. Hence BLAST scores between recently duplicated genes are, on average, larger than scores between orthologous genes. The top three tentative recent duplications for each member of the orthologue sets were examined and putative recent duplications were identified. Simultaneously, through a sequential, iterative process, gene orthologue sets were selected from largest to smallest BLAST scores making sure that every gene was contained in one, and only one, orthologue set.

Eight genomes were analysed to compute the global orthologue alignments. Organisms included in the analyses were chosen according to two criteria: they should span the diversity of life (see Supplementary Data for an analysis of 24 additional prokaryotic genomes) and their genomes should contain similar numbers of genes<sup>20</sup>. The slowly evolving yeast sequences facilitated comparisons with prokaryotic sequences (*Schizosaccharomyces pombe*, 5,000 genes, and *Saccharomyces cerevisiae*, 6,329). Both yeast genomes contained all nuclear genes, including those annotated as mitochondrial. The prokaryotes included four Bacteria (a  $\gamma$ -proteobacterium, *Xylella fastidiosa* 9a5c, 2,766 genes; an  $\alpha$ -proteobacterium, *Brucella melitensis* 16M, 3,198; a cyanobacterium, *Synechocystis* sp. PCC6803, 3,169; and a bacillus, *Staphylococcus aureus* MW2, 2,632) and four Archaea (a halobacterium, *Halobacterium* sp. NRC-1, 2,605; a methanococcus, *Methanosarcina mazei* Goe1, 3,371; an archaeoglobium, *Archaeoglobus fulgidus* DSM4304, 2,408; and an eocyte, *Sulfolobus tokodaii*, 2,826). *Archaeoglobus* and *Staphylococcus* were used as the conditioning genomes for Figs 1 and 2, respectively. The analyses shown in Fig. 1 are based on 2,408 orthologous gene sets. Of these, 433 sets contained all seven orthologues, whereas 239, 285, 292, 329, 293, 204 and 333 contained six, five, four, three, two, one and no orthologues, respectively. The analyses shown in Fig. 2 are based on 2,632 orthologous gene sets. Of these, 432 sets contained all seven orthologues, whereas 191, 208,

208, 230, 218, 266 and 879 contained six, five, four, three, two, one and no orthologues, respectively.

The big genome attraction artefact was partially compensated by using Paralinear/LogDet distances<sup>45,46</sup> for tree reconstruction, and the artefact was further reduced by using genomes of similar size; that is, containing similar numbers of genes. Bootstrappers Gambit<sup>47</sup> was used for the analyses as it, and possibly most quartet methods, can naturally accommodate graphs. Corrections for site-to-site variation used Pattern Filtering<sup>48</sup> and were applied to the gene order found in the conditioning genome, as this can considerably reduce errors introduced by violations of the assumption that sites are independent and identically distributed. The big genome attraction artefact did not appear to be significant for this set as the largest prokaryotic genomes were not directly connected to the eukaryotes in Figs 1 and 2.

Received 29 January; accepted 15 July 2004; doi:10.1038/nature02848.

- Lake, J. A., Sabatini, D. D. & Nonomura, Y. in *Ribosomes* (eds Nomura, M., Tissieres, A. & Lengyel, P.) 543–557 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1974).
- Woese, C. R. Archaeobacteria. *Sci. Am.* **244**, 98–105 (1981).
- Dayhoff, M. O. *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Silver Spring, Maryland, 1972).
- Pace, N. R., Olsen, G. J. & Woese, C. R. Ribosomal RNA phylogeny and the primary lines of evolutionary descent. *Cell* **45**, 325–326 (1986).
- Lake, J. A. Origin of the eukaryotic nucleus determined by rate-invariant analysis of ribosomal RNA sequences. *Nature* **331**, 184–186 (1988).
- Galtier, N., Tourasse, N. & Gouy, M. A nonhyperthermophilic common ancestor to extant life forms. *Science* **283**, 220–221 (1999).
- Gogarten, J. P. *et al.* Evolution of the vacuolar H<sup>+</sup>-ATPase—implications for the origin of eukaryotes. *Proc. Natl Acad. Sci. USA* **86**, 6661–6665 (1989).
- Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S. & Miyata, T. Evolutionary relationship of archaeobacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc. Natl Acad. Sci. USA* **86**, 9355–9359 (1989).
- Martin, W., Mustafa, A. Z., Henze, K. & Schnarrenberger, C. Higher-plant chloroplast and cytosolic fructose-1,6-bisphosphatase isoenzymes: Origins via duplication rather than prokaryote-eukaryote divergence. *Plant Mol. Biol.* **32**, 485–491 (1996).
- Brown, J. R. & Doolittle, W. F. Archaea and the prokaryote-to-eukaryote transition. *Microbiol. Mol. Biol. Rev.* **61**, 456–502 (1997).
- Feng, D. F., Cho, G. & Doolittle, R. F. Determining divergence times with a protein clock: Update and reevaluation. *Proc. Natl Acad. Sci. USA* **94**, 13028–13033 (1997).
- Gupta, R. S. Protein phylogenies and signature sequences: A reappraisal of evolutionary relationships among archaeobacteria, eubacteria, and eukaryotes. *Microbiol. Mol. Biol. Rev.* **62**, 1435–1491 (1998).
- Rivera, M. C., Jain, R., Moore, J. E. & Lake, J. A. Genomic evidence for two functionally distinct gene classes. *Proc. Natl Acad. Sci. USA* **95**, 6239–6244 (1998).
- Esser, C. *et al.* A genome phylogeny for mitochondria among  $\alpha$ -Proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes. *Mol. Biol. Evol.* doi:10.1093/molbev/msh160 (2004).
- Karlin, S., Mrazek, J. & Campbell, A. M. Compositional biases of bacterial genomes and evolutionary implications. *J. Bacteriol.* **179**, 3899–3913 (1997).
- Gogarten, J. P., Hilario, E. & Olendzenski, L. The tree of life. *ASM News* **63**, 404–405 (1997).
- Doolittle, W. F. Phylogenetic classification and the universal tree. *Science* **284**, 2124–2128 (1999).
- Campbell, A. M. Lateral gene transfer in prokaryotes. *Theor. Popul. Biol.* **57**, 71–77 (2000).
- Ochman, H. & Jones, I. B. Evolutionary dynamics of full genome content in *Escherichia coli*. *EMBO J.* **19**, 6637–6643 (2000).
- Lake, J. A. & Rivera, M. C. Deriving the genomic tree of life in the presence of horizontal gene transfer: Conditioned Reconstruction. *Mol. Biol. Evol.* **21**, 681–690 (2004).
- Dickerson, R. E. in *Diffraction and Related Studies* (ed. Srinivasan, R.) 227–249 (Pergamon, Oxford/New York, 1980).
- Snel, B., Bork, P. & Huynen, M. A. Genome phylogeny based on gene content. *Nature Genet.* **21**, 108–110 (1999).
- Fitz-Gibbon, S. T. & House, C. H. Whole genome-based phylogenetic analysis of free-living microorganisms. *Nucleic Acids Res.* **27**, 4218–4222 (1999).
- Tekaia, F., Lazcano, A. & Dujon, B. The genomic tree as revealed from whole proteome comparisons. *Genome Res.* **9**, 550–557 (1999).
- Montague, M. G. & Hutchison, C. A. Gene content phylogeny of herpesviruses. *Proc. Natl Acad. Sci. USA* **97**, 5334–5339 (2000).
- Lake, J. A., Henderson, E., Clark, M. W. & Matheson, A. T. Mapping evolution with ribosome structure: Intralineage constancy and interlineage variation. *Proc. Natl Acad. Sci. USA* **79**, 5948–5952 (1982).
- Gray, M. W., Burger, G. & Lang, B. F. Mitochondrial evolution. *Science* **283**, 1476–1481 (1999).
- Timmis, J. N., Ayliffe, M. A., Huang, C. Y. & Martin, W. Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. *Nature Rev. Genet.* **5**, 123–135 (2004).
- Gabaldon, T. & Huynen, M. A. Reconstruction of the proto mitochondrial metabolism. *Science* **301**, 609 (2003).
- Adams, K. L., Daley, D. O., Qiu, Y. L., Whelan, J. & Palmer, J. D. Repeated, recent and diverse transfers of a mitochondrial gene to the nucleus in flowering plants. *Nature* **408**, 354–357 (2000).
- Gray, M. W. Evolution of organellar genomes. *Curr. Opin. Genet. Dev.* **9**, 678–687 (1999).
- Collura, R. V. & Stewart, C. B. Insertions and duplications of mtDNA in the nuclear genomes of old-world monkeys and hominoids. *Nature* **378**, 485–489 (1995).
- Zischler, H., Geisert, H., vonHaseler, A. & Paabo, A. A nuclear fossil of the mitochondrial D-loop and the origin of modern humans. *Nature* **378**, 489–492 (1995).
- Margulis, L. *Origin of the Eukaryotic Cells* (Yale Univ. Press, New Haven, 1970).
- Gupta, R. S., Aitken, K., Falah, M. & Singh, B. Cloning of *Giardia lamblia* heat-shock protein Hsp70 homologs—Implications regarding origin of eukaryotic cells and of endoplasmic-reticulum. *Proc. Natl Acad. Sci. USA* **91**, 2895–2899 (1994).
- Martin, W. & Muller, M. The hydrogen hypothesis for the first eukaryote. *Nature* **392**, 37–41 (1998).
- Lake, J. A. & Rivera, M. C. Was the nucleus the 1st endosymbiont. *Proc. Natl Acad. Sci. USA* **91**, 2880–2881 (1994).
- Moreira, D. & Lopez-Garcia, P. Symbiosis between methanogenic archaea and  $\delta$ -proteobacteria as the origin of eukaryotes: The syntrophic hypothesis. *J. Mol. Evol.* **47**, 517–530 (1998).
- Horiike, T., Hamada, K., Kanaya, S. & Shinozawa, T. Origin of eukaryotic cell nuclei by symbiosis of Archaea in Bacteria is revealed by homology-hit analysis. *Nature Cell Biol.* **3**, 210–214 (2001).
- Rivera, M. C. & Lake, J. A. Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. *Science* **257**, 74–76 (1992).
- Daubin, V., Gouy, M. & Perriere, B. A phylogenomic approach to bacterial phylogeny: Evidence of a core of genes sharing a common history. *Genome Res.* **12**, 1080–1090 (2002).
- Brochier, C., Forterre, P. & Gribaldo, S. Archaeal phylogeny based on proteins of the transcription and translation machineries: tackling the *Methanopyrus kandleri* paradox. *Genome Biol.* **5**, R17 (2004).
- Wolf, Y. I., Rogozin, I. B., Grishin, N. V., Tatusov, R. L. & Koonin, E. V. Genome trees constructed using five different approaches suggest new major bacterial clades. *BMC Evol. Biol.* **1**, 8 (2001).
- Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
- Lake, J. A. Reconstructing evolutionary trees from DNA and protein sequences—Paralinear distances. *Proc. Natl Acad. Sci. USA* **91**, 1455–1459 (1994).
- Lockhart, P. J., Steel, M. A., Hendy, M. D. & Penny, D. Recovering evolutionary trees under a more realistic model of sequence evolution. *Mol. Biol. Evol.* **11**, 605–612 (1994).
- Lake, J. A. Calculating the probability of multitaxon evolutionary trees—Bootstrappers Gambit. *Proc. Natl Acad. Sci. USA* **92**, 9662–9666 (1995).
- Lake, J. A. Optimally recovering rate variation information from genomes and sequences: Pattern filtering. *Mol. Biol. Evol.* **15**, 1224–1231 (1998).
- Woese, C. R., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria and Eucarya. *Proc. Natl Acad. Sci. USA* **87**, 4576–4579 (1990).

**Supplementary Information** accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank A. B. Simonson for helpful discussions and editing, M. Kowalczyk for illustrations, J. Garey for suggesting the original title ‘One ring to rule them all’, and J. A. Servin and R. G. Skophammer for suggestions. This work was supported by grants from the National Science Foundation, NASA Astrobiology Program, The Department of Energy, and the National Institutes of Health to J.A.L.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to J.A.L. (lake@mbi.ucla.edu).



# Implications for hydrologic processes on Mars from extensive bedrock outcrops throughout Terra Meridiani

Brian M. Hynek

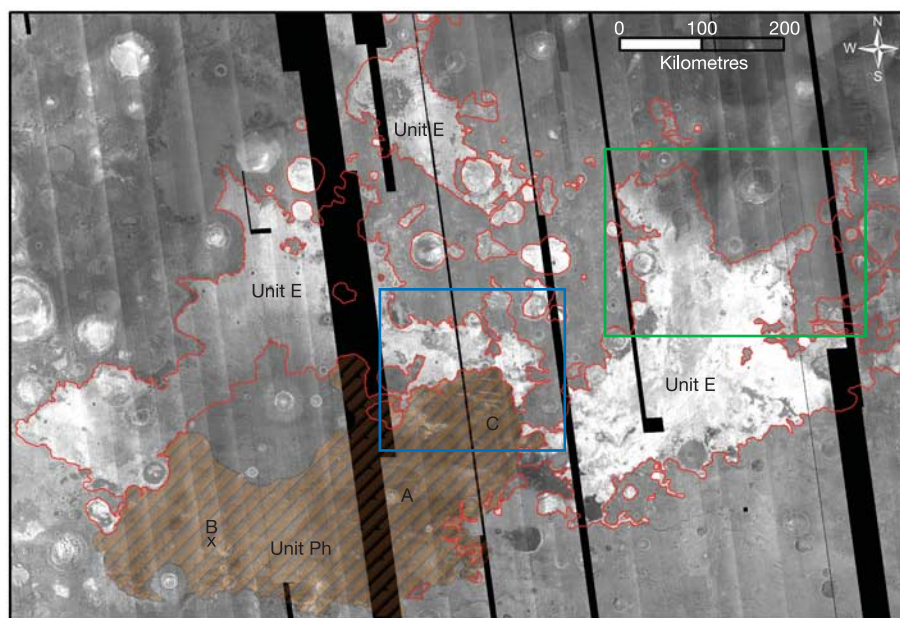
Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder, Colorado 80309, USA

Grey haematite was recently detected in the Terra Meridiani region of Mars<sup>1,2</sup> by the Thermal Emission Spectrometer onboard the Mars Global Surveyor spacecraft. The formation of haematite on Earth often requires liquid water to be present for long periods of time, making this an important target for deciphering the history of water on Mars. The Mars Exploration Rover Opportunity landed in Meridiani early in 2004 and has since discovered light-toned bedrock outcrops rich in chemical and textural signatures of long-term water interaction locally at the landing site<sup>3</sup>. Here I use remote sensing data to show that the light-toned outcrops at the landing site are not a local phenomenon. Instead, they are observable throughout the haematite-bearing plains in both visible and thermal infrared remote sensing data. Moreover, the light-toned material outcrops around much of the margin and is mappable for hundreds of kilometres to the north, east and west of the plains. I infer that  $3 \times 10^5 \text{ km}^2$  of this material is exposed over  $20^\circ$  of longitude, indicating the

extended presence of surface or near-surface water over a large region of Mars.

Previous geomorphic mapping of the Meridiani region indicated that the haematite-rich plains unit (herein referred to as unit Ph, for 'plains, haematite') is an upper stratum in a thick (hundreds of metres) stack of layered deposits<sup>4</sup> that date from the Noachian epoch ( $>3.7$  Gyr ago)<sup>4,5</sup>. A major stratigraphic component of these layered deposits are high albedo, high thermal inertia bedforms<sup>4,6,7</sup>. Thermal inertia, determined from remotely sensed observations, is used to infer near-surface particle size, degree of induration, rock abundance and bedrock exposure<sup>8,9</sup>. The high inertia and albedo layers were called 'etched' terrain (herein termed unit E) because of their rough appearance resulting from aeons of differential erosion<sup>4</sup>. For simplicity, unit E will be referred to as a single unit, even though many layers are visible, varying in competency, thickness and thermophysical properties.

The Mars Exploration Rover (MER) science team noted that the sulphur-rich bedrock discovered in the walls of the small crater where Opportunity landed is layered at the millimetre scale and that haematite-rich spherules are weathering out from the laminae<sup>3</sup>. The morphologic and geochemical evidence strongly suggest aqueous alteration, possibly in a standing body of water. Here it is argued that the outcrops observed by Opportunity are the upper layers of unit E, and that these are correlative with exposures throughout unit Ph and well beyond. Prior to landing, unit E was thought to underlie part of unit Ph and it outcropped around much of the margin to the east, north and west<sup>4,7</sup>. In this study, I have used Thermal Emission Imaging System (THEMIS) data to expand and revise previous mapping efforts as well as characterize the thermophysical properties of the unit (Fig. 1). The average THEMIS-derived thermal



**Figure 1** THEMIS-derived thermal inertia map of the Terra Meridiani region of Mars<sup>16</sup>. Low values (dark) correspond to unconsolidated fine grains and higher values (light) represent mixed surfaces with increasing induration and rock abundance. There is nearly complete THEMIS night-time coverage at a resolution of 100 m per pixel. Relative to the Thermal Emission Spectrometer (TES) data, THEMIS-derived thermal inertias offer much higher spatial resolution but their values are less accurate for several reasons<sup>16,17</sup>. For this study, each THEMIS thermal inertia image was calibrated to the TES data covering the same geographic extent to minimize these uncertainties. The images were then mosaicked together for a relatively seamless data map. Geomorphic units of E and Ph have been overlaid on the mosaic. The boundary of unit Ph was refined from earlier work<sup>4</sup>

where the unit was originally identified from a unique TES spectral signature<sup>2</sup>. Unit E was identified and its extent was mapped with images from the MOC WA and MOC NA and the THEMIS instrument. Collectively, the images have resolutions ranging from 2 to 250 m and cover the visible and thermal infrared portions of the spectrum. The THEMIS-derived thermal inertia data were also crucial in mapping the high-inertia unit E. The etched material outcrops around much of the low thermal inertia unit Ph, and extends well beyond it. Boxed regions correspond to Fig. 2 (blue) and Fig. 3 (green). Labels A, B and C indicate the locations of images in Fig. 4a–c. The 'x' marks the location of the Opportunity Rover. Geographic extents:  $5^\circ \text{S}$ – $10^\circ \text{N}$ ,  $12^\circ \text{W}$ – $10^\circ \text{E}$ .

inertia for unit E is relatively high,  $\sim 360 \text{ J m}^{-2} \text{ K}^{-1} \text{ s}^{-1/2}$ , strengthening the argument that the unit contains substantial bedrock. Contrastingly, unit Ph has an average value of  $\sim 175 \text{ J m}^{-2} \text{ K}^{-1} \text{ s}^{-1/2}$  and the surface is inferred to be primarily fine-grained, unconsolidated material. Interestingly, there is great variability in thermal inertia (up to  $80 \text{ J m}^{-2} \text{ K}^{-1} \text{ s}^{-1/2}$ ) between subjacent layers within unit E, which is evidence of substantial compositional or physical differences. In this region of Mars, the thermophysical properties correlate well with previously defined geomorphic units, indicating local control of physical properties and composition, and making these data useful for mapping the terrain (Fig. 2).

Unit E has a surface exposure covering  $2.3 \times 10^5 \text{ km}^2$  in the Meridiani region (Fig. 1) and large, contiguous outcrops are exposed laterally and vertically. For example, a deep trough just north of unit Ph cuts through  $\sim 500 \text{ m}$  of stratigraphy and unit E is exposed across the entire section. Layering is evident throughout the unit and in places individual layers can be continuously mapped for  $>250 \text{ km}$ , particularly to the east of unit Ph. These layers are often extremely flat and maintain a nearly constant elevation over great distances.

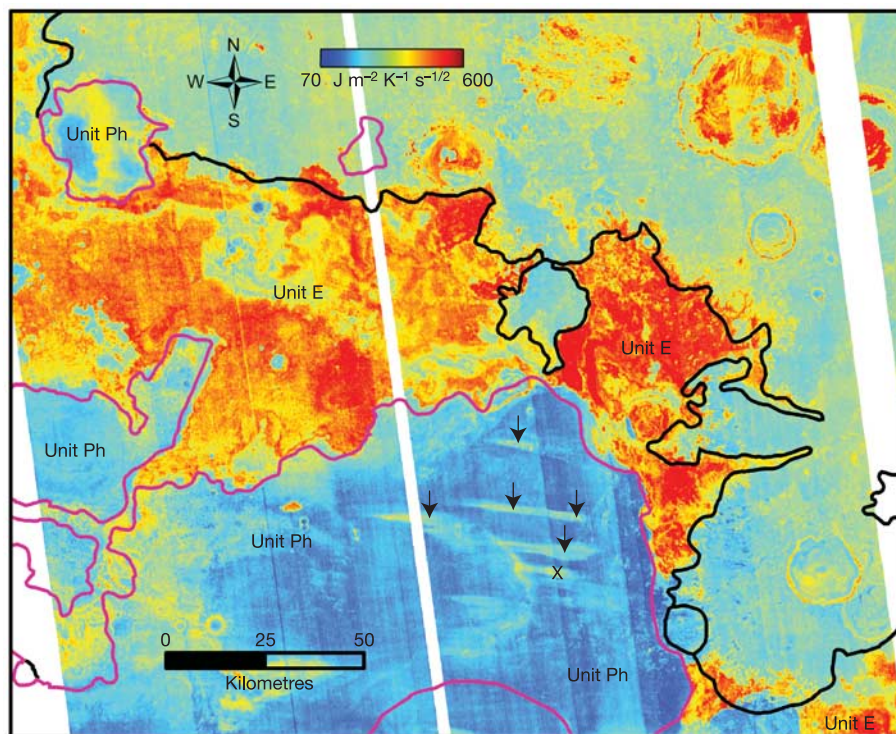
In numerous places, the northern edge of unit E disappears under younger units (Fig. 3). Disjointed pieces of unit E are seen and become increasingly obscured with distance from the primary exposures. This is most evident to the northeast, where a dust mantle is thought to cover the region<sup>9,10</sup>. The isolated, generally high thermal inertia, layered units seen throughout western Arabia Terra to the north of the Meridiani region<sup>6,11–13</sup> may be stratigraphically related to unit E, but more work is required to establish any connection.

Recent data from the Mars Orbiter Camera (MOC) and THEMIS

show that unit E is exposed throughout unit Ph in several forms: (1) as layers within crater walls, (2) mixed within crater ejecta, (3) exposed on the plains as disjointed outcrops, and (4) within streaks across dune fields (Fig. 4). The first three forms occur across the entire unit Ph and the last is predominantly seen on the northeastern lobe of the unit. In THEMIS data, layers within crater rims have relatively high thermal inertias, often several hundred units above the adjacent haematite-bearing plains. The layering extends down hundreds of metres in the larger impacts. However, even small impact craters, with diameters  $\ll 1 \text{ km}$  and implied depths of a few to tens of metres have pierced the underlying bright terrain (Fig. 4b). In fact, nearly every fresh crater throughout unit Ph (that is, those which formed upon the pre-existing dark plain and underlying stratigraphy) has light-toned layered materials composing its walls. In addition to layering within the walls, nearly all the fresh impact craters on unit Ph have bright, high-thermal-inertia material within their crater ejecta that is interpreted to be excavated material from unit E (Fig. 4a).

Orbiter observations like these led to the pre-landing hypothesis that the Opportunity Rover may have a chance to sample this underlying bright terrain (unit E), particularly if impact craters could be investigated<sup>7,14</sup>. This prediction was confirmed with the Rover's identification of light-toned layered outcrops in a  $\sim 20\text{-m}$ -diameter crater<sup>3</sup>. The occurrence of bedrock in a crater that is only a few metres deep indicates that the dark unit Ph is a very thin deposit on top of the much thicker unit E. Similar exposures of unit E across the entire haematite plain suggest that a thin unit Ph resting on a much thicker unit E may be commonplace.

Earlier work recognized localized minor occurrences of exposed unit E within unit Ph<sup>4,7</sup>. A large increase in the MOC narrow-angle



**Figure 2** THEMIS-derived thermal inertia map indexed on Fig. 1 overlaid with geomorphic unit boundaries from ref. 4. (Only units Ph and E are included, for simplicity.) The unit boundaries were defined strictly on geomorphic characteristics from visible images and topography, with no knowledge of fine-scale thermal inertia. There is a clear correlation, however, between high-inertia areas and unit E and between low-inertia regions and unit Ph. The contact is very sharp, with thermal inertia values increasing by several hundred

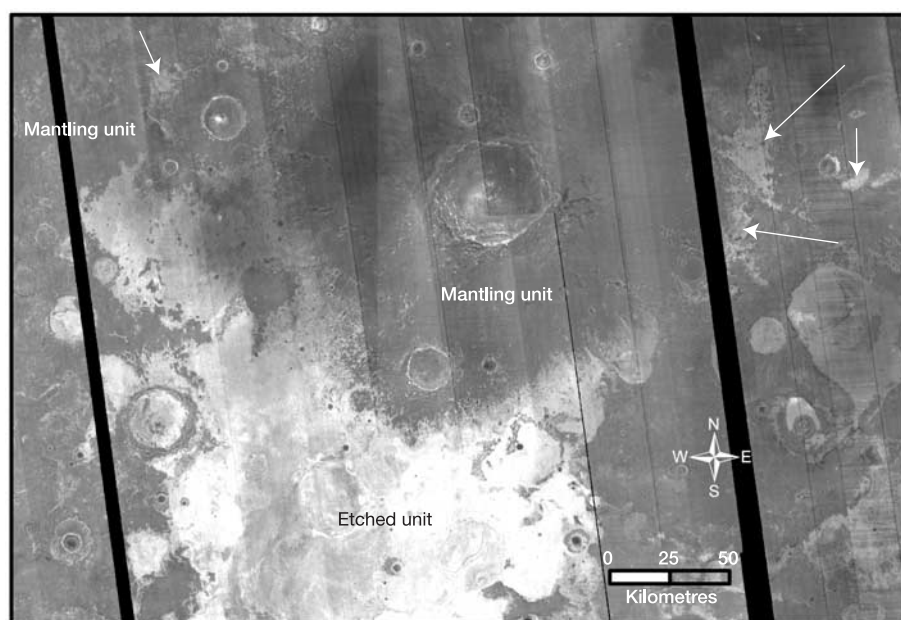
units over  $<0.5 \text{ km}$  when crossing from unit Ph to E. A similar correlation between these geomorphic units and TES thermal inertia was reported in ref. 7; however, with an increase of more than an order of magnitude in spatial resolution it is remarkable that the thermophysical properties still reflect the geomorphology so closely. Numerous wind scour streaks are seen on unit Ph (arrows) and are described in the text. The 'X' marks the location of Fig. 4c.



camera (MOC NA) and THEMIS image coverage, however, has recently shown that exposures of unit E occur throughout most of unit Ph as degraded and disjointed outcrops (see, for example, Fig. 4b). Areas with a higher concentration of these bright features, with either positive or negative relief, have a mottled appearance in MOC images and THEMIS thermal inertia maps (Fig. 4b). The higher thermal inertia values in these areas are similar to those of bright layers within the upper walls of nearby craters. The low thermal inertia values correspond to adjacent dark portions of unit Ph. These observations suggest that the bright, high thermal inertia areas in the mottled terrain correlate with unit E exposures in proximal craters and beyond unit Ph. The dark areas are hypothesized to be 'typical' dark, haematite-bearing plains material. The

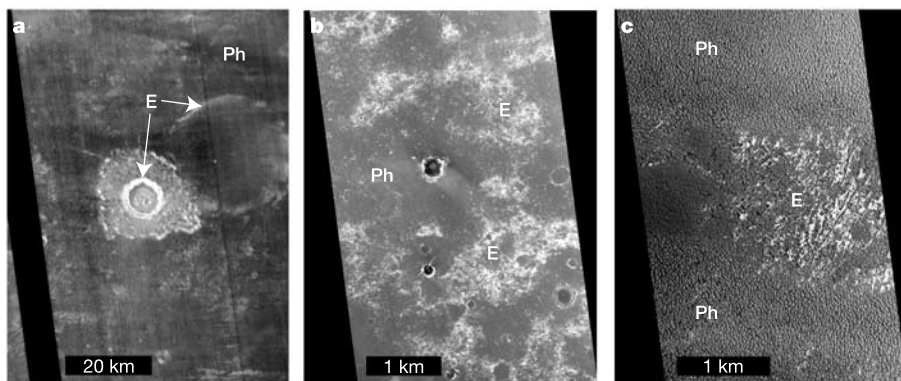
mottled terrain is evident as patches across most of unit Ph, including directly north of the Opportunity landing site, as seen in Fig. 4b.

The final type of exposure of unit E within unit Ph is manifest in relatively high thermal inertia streaks across the dark plains, as evident in THEMIS data (arrows in Fig. 2; detailed view shown in Fig. 4c). These streaks occur almost exclusively on the northeastern portion of unit Ph, which has a very different appearance in comparison to the rest of the unit. MOC and THEMIS images show an extensive dune field covering the entire northeast lobe of unit Ph. This region has the lowest THEMIS-derived thermal inertia values of the unit, averaging  $\sim 100 \text{ J m}^{-2} \text{ K}^{-1} \text{ s}^{-1/2}$ . Relatively high thermal inertia streaks stretching from east to west across the dune



**Figure 3** Greyscale mosaic of THEMIS-derived thermal inertia indexed on Fig. 1. Dark areas are expected to be unconsolidated fine-grained surfaces whereas bright areas probably contain rock outcrops associated with the etched unit. Younger mantling units in

the north part of the image obscure unit E. Outliers are exposed in places (arrows), indicating that the etched unit continues to the north until it is entirely buried, leaving its full extent unknown.



**Figure 4** Examples of etched terrain (E) within the haematite-bearing plains (Ph). **a**, THEMIS-derived thermal inertia mosaic showing etched layers within the rims of a superposed crater and a larger, mostly buried impact basin. These layers begin just below the rim of the impact crater. The ejecta of the fresh crater is evident where high thermal inertias result from excavation of the underlying high-inertia material and alteration due to the impact process. These thermal inertia values are typically higher than those of the subjacent plains but lower than those of most etched layers within the crater walls. **b**, MOC NA image E10-00916 of mottled terrain just north of the Opportunity landing site,

composed of light-toned outcrops (unit E) embayed by dark mobile material (unit Ph). Layers in unit E show variable competency and differential erosion by aeolian processes that have resulted in cliffs, mesas, troughs and slopes. **c**, MOC NA image E11-03974 showing the etched unit exposed within a dark, haematite-bearing dune field that corresponds with a high inertia streak on unit Ph (location marked with 'X' in Fig. 2). These features have dimensions up to 6 km wide and 35 km long and occur exclusively on the northeast portion of unit Ph. The locations of the images in **a** to **c** are shown in Fig. 1.



field are seen in the THEMIS data (Fig. 2), yet are absent or barely discernible in the MOC wide-angle camera (MOC WA) and Mars Orbiter Laser Altimeter topographic data. One MOC NA image crosses several of these features and reveals that they are exposures of underlying bright terrain where dunes are not present (Fig. 4c). These features are probably locations of aeolian scour, which has exposed unit E from beneath the pervasive dune field.

The ubiquitous exposure of unit E within unit Ph strongly suggests that it underlies all of the haematite-rich plains. Unit E is exposed in craters of all sizes, down to the resolution of the MOC NA images ( $\sim 2$  m), indicating that unit Ph is very thin. In all cases, unit E appears as bright, high thermal inertia outcrops that contrast sharply with unit Ph. Evidence of expansive dune fields and embayment of positive topographic expressions of unit E suggests that the thin, dark, haematite-bearing layer is mobile and may be the primary soil constituent. Unit Ph is probably a lag deposit made of dense, coarse, haematite-rich grains derived from the erosion of previously existing overlying layers, and this is the source of the haematite signature detected by the TES instrument. Conversely, unit E consists of in-place underlying bedrock, as is evident from its high thermal inertia and its layered exposures seen in images.

In the Opportunity landing site crater,  $\sim 1/2$ -m-thick, finely layered units are exposed in the rim<sup>3</sup> and embayed by dark, haematite-rich soil<sup>15</sup>. These outcrops are probably the uppermost expressions of unit E and correlate with light-toned layers observed throughout the haematite-bearing plain, an area of  $1 \times 10^5$  km<sup>2</sup>. Moreover, unit E outcrops along roughly half of the unit Ph margin and is mappable over hundreds of kilometres to the east, north and west. In places, unit E disappears underneath younger mantling materials, masking its full extent. In totality, exposures of this etched material cover an area of  $3.3 \times 10^5$  km<sup>2</sup> that spans  $20^\circ$  of longitude and  $14^\circ$  of latitude, more than three times the extent of the haematite-bearing plain itself. Opportunity has detected high concentrations of sulphur-bearing minerals and unique textures in the observed bedrock that are indicative of water saturation for long periods of time<sup>3</sup>. Thus, whatever aqueous process altered, and perhaps formed, the layered units at the landing site<sup>3</sup> must have acted over an extensive area and a large volume of material. If the outcrops at the landing site are indeed a result of deposition in a sea, then that body of water had to be comparable in area to Earth's Baltic Sea and must have been deep enough, and persisted long enough, to build up at least a 0.5-km-thick stack of sediments. For this to occur, the ancient global climate of Mars must have been very different from its present climate and have lasted for an extended period. □

Received 20 April; accepted 28 July 2004; doi:10.1038/nature02902.

- Christensen, P. R. *et al.* Detection of crystalline hematite mineralization on Mars by the Thermal Emission Spectrometer: Evidence for near-surface water. *J. Geophys. Res.* **105**, 9623–9642 (2000).
- Christensen, P. R., Morris, R. V., Lane, M. D., Bandfield, J. L. & Malin, M. C. Global mapping of martian hematite mineral deposits: Remnants of water-driven processes on early Mars. *J. Geophys. Res.* **106**, 23873–23886 (2001).
- Squyres, S. W. *et al.* Initial results from the MER ATHENA science investigation at Gusev Crater and Meridiani Planum. *Lunar Planet. Sci. Conf. XXXV*, 2187 (2004).
- Hynek, B. M., Arvidson, R. E. & Phillips, R. J. Geologic setting and origin of Terra Meridiani hematite deposit on Mars. *J. Geophys. Res.* **107**, doi:10.1029/2002JE001891 (2002).
- Lane, M. D., Christensen, P. R. & Hartmann, W. K. Utilization of the THEMIS visible and infrared imaging data for crater population studies of the Meridiani Planum landing site. *Geophys. Res. Lett.* **30**, doi:10.1029/2003GL017183 (2003).
- Edgett, K. S. & Malin, M. C. Martian sedimentary rock stratigraphy: Outcrops and interbedded craters of northwest Sinus Meridiani and southwest Arabia Terra. *Geophys. Res. Lett.* **29**, doi:10.1029/2002GL016515 (2002).
- Arvidson, R. E. *et al.* Mantled and exhumed terrains in Terra Meridiani, Mars. *J. Geophys. Res.* **108**, doi:10.1029/2002JE001982 (2003).
- Palluconi, F. D. & Kieffer, H. H. Thermal inertia mapping of Mars from 60 degrees S to 60 degrees N. *Icarus* **45**, 415–426 (1981).
- Mellon, M. T., Jakosky, B. M., Kieffer, H. H. & Christensen, P. R. High-resolution thermal inertia mapping from the Mars Global Surveyor Thermal Emission Spectrometer. *Icarus* **148**, 437–455 (2000).
- Tanaka, K. L. Dust and ice deposition in the martian geologic record. *Icarus* **144**, 254–266 (2000).
- Malin, M. C. & Edgett, K. S. Sedimentary rocks of early Mars. *Science* **290**, 1927–1937 (2000).
- Malin, M. C. & Edgett, K. S. Mars Global Surveyor Mars Orbiter Camera: Interplanetary cruise through primary mission. *J. Geophys. Res.* **106**, 23429–23570 (2001).

- Hynek, B. M., Phillips, R. J. & Arvidson, R. E. Explosive volcanism in the Tharsis region: Global evidence in the martian geologic record. *J. Geophys. Res.* **108**, doi:10.1029/2003JE002062 (2003).
- Hynek, B. M. *et al.* Thermophysical properties of Meridiani Planum, Mars. *Lunar Planet. Sci. Conf. XXXV*, 1899 (2004).
- Christensen, P. R. *et al.* Mini-TES observations of the Gusev and Meridiani landing sites. *Lunar Planet. Sci. Conf. XXXV*, 2186 (2004).
- Putzig, N. E. *et al.* Mars thermal inertia from THEMIS data. *Lunar Planet. Sci. Conf. XXXV*, 1863 (2004).
- Jakosky, B. M. *et al.* Remote-sensing of the thermophysical properties of the MER and Beagle II landing sites on Mars. *J. Geophys. Res.* (submitted) (2004).

**Acknowledgements** I thank N. E. Putzig and M. T. Mellon for THEMIS-derived thermal inertia products used in this study and B. M. Jakosky and J. R. Zimbelman for comments on this manuscript. This work was supported by the National Aeronautics and Space Agency (NASA).

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to B.M.H. (brian.hynek@lasp.colorado.edu).

## Coherent dynamics of a flux qubit coupled to a harmonic oscillator

I. Chiorescu<sup>1\*</sup>, P. Bertet<sup>1</sup>, K. Semba<sup>1,2</sup>, Y. Nakamura<sup>1,3</sup>, C. J. P. M. Harmans<sup>1</sup> & J. E. Mooij<sup>1</sup>

<sup>1</sup>Quantum Transport group, Kavli Institute of NanoScience, Delft University of Technology and Delft Institute for Micro Electronics and Submicron Technology (DIMES), Lorentzweg 1, 2628 CJ, Delft, The Netherlands

<sup>2</sup>NTT Basic Research Laboratories, NTT Corporation, 3-1 Morinosato-Wakamiya, Atsugi 243-0198, Japan

<sup>3</sup>NEC Fundamental Research Laboratories, 34 Miyukigaoka, Tsukuba, Ibaraki 305-8501, Japan

\* Present address: Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824, USA

In the emerging field of quantum computation<sup>1</sup> and quantum information, superconducting devices are promising candidates for the implementation of solid-state quantum bits (qubits). Single-qubit operations<sup>2–6</sup>, direct coupling between two qubits<sup>7–10</sup> and the realization of a quantum gate<sup>11</sup> have been reported. However, complex manipulation of entangled states—such as the coupling of a two-level system to a quantum harmonic oscillator, as demonstrated in ion/atom-trap experiments<sup>12,13</sup> and cavity quantum electrodynamics<sup>14</sup>—has yet to be achieved for superconducting devices. Here we demonstrate entanglement between a superconducting flux qubit (a two-level system) and a superconducting quantum interference device (SQUID). The latter provides the measurement system for detecting the quantum states; it is also an effective inductance that, in parallel with an external shunt capacitance, acts as a harmonic oscillator. We achieve generation and control of the entangled state by performing microwave spectroscopy and detecting the resultant Rabi oscillations of the coupled system.

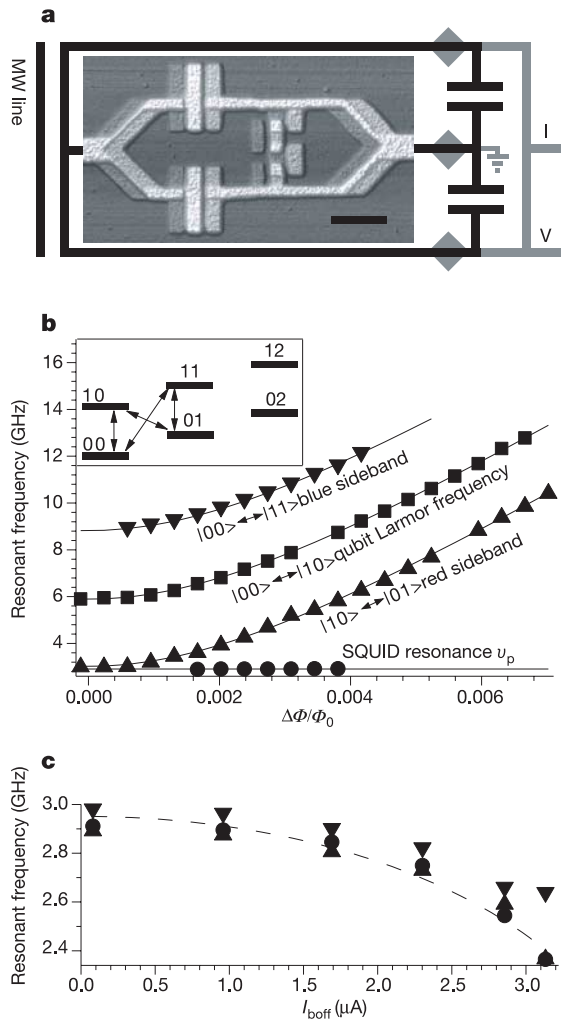
The device was realized by electron-beam lithography and metal evaporation. The qubit–SQUID geometry is shown in Fig. 1a: a large loop interrupted by two Josephson junctions (the SQUID) is merged with the smaller loop on the right-hand side comprising three in-line Josephson junctions (the flux qubit)<sup>15</sup>. By applying a perpendicular external magnetic field, the qubit is biased around  $\Phi_0/2$ , where  $\Phi_0 = h/2e$  is the flux quantum. Previous spectroscopy<sup>16</sup> and coherent time-domain experiments<sup>6</sup> have shown that the flux qubit is a controllable two-level system with ‘spin-up/spin-down’ states corresponding to persistent currents flowing in ‘clockwise/

anticlockwise' directions and coupled by tunnelling. Here we show that a stronger qubit–SQUID coupling allows us to investigate the coupled dynamics of a 'qubit–harmonic oscillator' system.

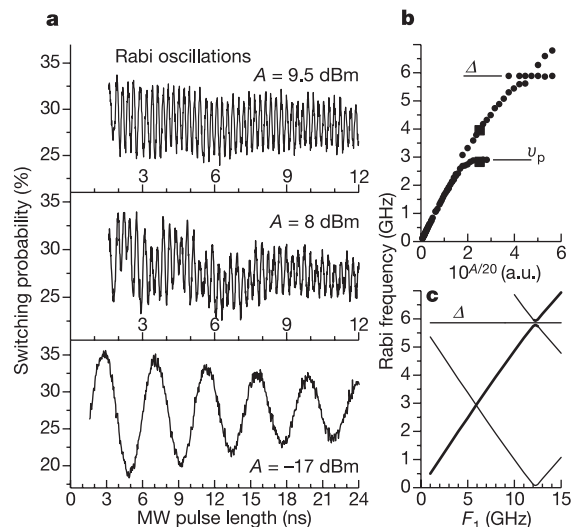
The qubit hamiltonian is defined by the charging and Josephson energy of the qubit outer junctions ( $E_C = e^2/2C$  and  $E_J = \hbar I_C/4\pi e$ , where  $C$  and  $I_C$  are their capacitance and critical current)<sup>16</sup>. In a

two-level truncation, the hamiltonian becomes  $H_q/\hbar = -\varepsilon\sigma_z/2 - \Delta\sigma_x/2$ , where  $\sigma_{z,x}$  are the Pauli matrices in the spin-up/spin-down basis,  $\Delta$  is the tunnel splitting and  $\varepsilon \approx I_p\Phi_0(\gamma_q - \pi)/\hbar\pi$  ( $I_p$  is the qubit maximum persistent current and  $\gamma_q$  is the superconductor phase across the three junctions). The resulting energy level spacing represents the qubit Larmor frequency  $F_L = \sqrt{\Delta^2 + \varepsilon^2}$ . The SQUID dynamics is characterized by the Josephson inductance of the junctions  $L_J \approx 80$  pH, shunt capacitance  $C_{sh} \approx 12$  pF (see Fig. 1a) and self-inductances  $L_{sl} \approx 170$  pH of the SQUID and shunt-lines. In our experiments, the SQUID circuit behaves like a harmonic oscillator described by  $H_{sq} = \hbar\nu_p(a^\dagger a + 1/2)$ , where  $2\pi\nu_p = 1/\sqrt{(L_J + L_{sl})C_{sh}}$  is called the plasma frequency and  $a$  ( $a^\dagger$ ) is the plasmon annihilation (creation) operator. Henceforth  $|\beta n\rangle$  represents the state with the qubit in the ground ( $\beta = 0$ ) or excited ( $\beta = 1$ ) level, and the oscillator on the  $n$ th level ( $n = 0, 1, 2, \dots$ ). The corresponding level diagram is sketched in Fig. 1b (inset). The coupling between the qubit and the oscillator originates from the current distribution in the shared branches (Fig. 1a), and gives rise to an interaction hamiltonian  $H_{q-sq} = \lambda\sigma_z(a + a^\dagger)$  with  $\lambda \approx 0.2$  GHz in our device<sup>17</sup> (the estimated qubit–SQUID coupling is  $M \approx 20$  pH).

Measurements are performed at  $T = 25$  mK using low-noise circuitry to minimize decoherence, relaxation and thermal activation. The system is first initialized by allowing it to relax to the  $|00\rangle$  ground state. With successive resonant microwave pulses we achieve controlled superposition of various  $|\beta n\rangle$  states, as shown below. The readout<sup>6</sup> is performed by applying a short current pulse  $I_b$  ( $\sim 30$  ns) and by monitoring whether the SQUID switches to the finite-voltage state. After averaging typically 10,000 readouts, we obtain the probability  $P_{sw}(I_b)$  which for properly chosen parameters is proportional to the excited state occupancy. In the



**Figure 1** Qubit–SQUID device and spectroscopy. **a**, Atomic force micrograph of the SQUID (large loop) merged with the flux qubit (the smallest loop closed by three junctions); the qubit to SQUID area ratio is 0.37. Scale bar, 1  $\mu$ m. The SQUID (qubit) junctions have a critical current of 4.2 (0.45)  $\mu$ A. The device is made of aluminium by two symmetrically angled evaporations with an oxidation step in between. The surrounding circuit shows aluminium shunt capacitors and lines (in black) and gold quasiparticle traps<sup>3</sup> and resistive leads (in grey). The microwave field is provided by the shortcut of a coplanar waveguide (MW line) and couples inductively to the qubit. The current line (I) delivers the readout pulses, and the switching event is detected on the voltage line (V). **b**, Resonant frequencies indicated by peaks in the SQUID switching probability when a long microwave pulse excites the system before the readout pulse. Data are represented as a function of the external flux through the qubit area away from the qubit symmetry point. Inset, energy levels of the qubit–oscillator system for some given bias point. The blue and red sidebands are shown by down- and up-triangles, respectively; continuous lines are obtained by adding 2.96 GHz and  $-2.90$  GHz, respectively, to the central continuous line (numerical fit). These values are close to the oscillator resonance  $\nu_p$  at 2.91 GHz (solid circles), and we attribute the small differences to the slight dependence of  $\nu_p$  on qubit state. **c**, The plasma resonance (circles) and the distances between the qubit peak (here  $F_L = 6.4$  GHz) and the red/blue (up/down triangles) sidebands as a function of an offset current  $I_{boff}$  through the SQUID. The data are close to each other and agree well with the theoretical prediction for  $\nu_p$  versus offset current (dashed line).



**Figure 2** Rabi oscillations at the qubit symmetry point  $\Delta = 5.9$  GHz. **a**, Switching probability as a function of the microwave pulse length for three microwave nominal powers; decay times are of the order of 25 ns. For  $A = 8$  dBm, bi-modal beatings are visible (the corresponding frequencies are shown by the filled squares in **b**). **b**, Rabi frequency, obtained by fast Fourier transformation of the corresponding oscillations, versus microwave amplitude. In the weak driving regime, the linear dependence is in agreement with estimations based on sample design. A first splitting appears when the Rabi frequency is  $\sim \nu_p$ . In the strong driving regime, the power independent Larmor precession at frequency  $\Delta$  gives rise to a second splitting. **c**, This last aspect is obtained in numerical simulations where the microwave driving is represented by a term  $(1/2)\hbar F_1 \cos(\Delta t)$  and a small deviation from the symmetry point (100 MHz) is introduced in the strong driving regime (the thick line indicates the main Fourier peaks). Radiative shifts<sup>20</sup> at high microwave power could account for such a shift in the experiment.

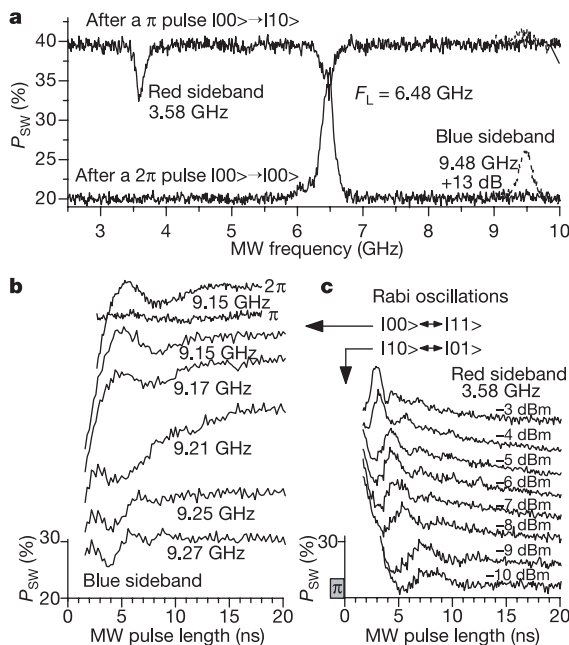
following, we first show the spectroscopy of the coupled qubit–oscillator system and Rabi oscillations of the qubit. Next we demonstrate coherent dynamics of the coupled system.

We performed spectroscopy of the coupled qubit–oscillator system by applying a long (300 ns) microwave pulse with various frequencies and measuring the SQUID switching probability. Peaks and dips are observed and their resonant frequencies as a function of  $\Delta\Phi = \Phi_{\text{ext}} - \Phi_0/2$  are given in Fig. 1b. We obtain one manifold of three resonances spaced by  $\sim 2.91$  GHz. This frequency coincides with the designed oscillator eigenfrequency  $\nu_p$ . In addition, we observe a spectroscopic peak or dip that depends only weakly on the magnetic field (circles in Fig. 1b). For lower microwave power, only the qubit band (squares) remains visible. A numerical fit (continuous line) of this band leads to  $E_J = 225$  GHz,  $E_C = 7.3$  GHz, and the ratio of area of qubit junctions  $\alpha = 0.76$  ( $\Delta = 5.9$  GHz,  $I_p = 275$  nA). The appearance of the manifold instead of a single resonance is due to the qubit coupling with the oscillator mode  $\nu_p$  (ref. 18). Similarly to atomic physics, we call the  $|00\rangle$  to  $|11\rangle$  ( $|01\rangle$  to  $|10\rangle$ ) transitions the blue (red) sidebands (see the ladder energy diagram of the  $|\beta n\rangle$  states in Fig. 1b inset). We note that near the qubit symmetry point, the closeness of the oscillator resonance and the red sideband, visible owing to a small thermal occupation of the  $|01\rangle$  state, is purely accidental. To verify that

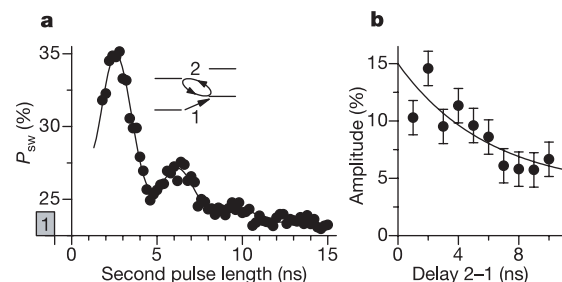
the oscillator involved is indeed the SQUID plasma mode, we repeated the above measurements in the presence of an offset bias current  $I_{\text{boff}}$  which decreases the plasma frequency following<sup>19</sup>  $\nu_p(1 - (I_{\text{boff}}/I_c)^2)^{1/4}$ , where  $I_c$  is the SQUID critical current (4.2  $\mu$ A). The data in Fig. 1c show the distance between the qubit peak for  $F_L = 6.4$  GHz and the blue/red sidebands (down/up triangles) that decreases together with the oscillator resonance (circles).

To realize quantum operations on the qubit only, we apply a resonant microwave pulse with frequency  $F_{\text{mw}} = F_L$ . The operation is performed at the qubit symmetry point  $\gamma_q = \pi$  where  $F_L = \Delta$ . In Fig. 2a, the SQUID switching probability is plotted against the microwave pulse length for three microwave power levels. The observed Rabi oscillations decay within  $\sim 30$  ns. Remarkably, we can reach Rabi frequencies comparable to the Larmor frequency (up to 6.6 GHz). Using Fourier transformation, we extract the Rabi frequency as a function of the microwave amplitude (Fig. 2b). In the weak driving regime, the Rabi frequency increases linearly with the microwave power, as expected<sup>6</sup>. Near the oscillator resonance  $\nu_p$ , we see two frequencies in the spectrum, a behaviour which is probably caused by the qubit–oscillator coupling. At even higher microwave powers, the spectrum exhibits again a second frequency component at  $\Delta$ . A qualitatively similar behaviour is also obtained in numerical simulations (see Fig. 2c) when we consider the qubit driven by an additional term  $(1/2)\hbar F_1 \cos(\Delta t)$  in  $H_q$  ( $F_1$  and  $\Delta$  are the microwave amplitude and frequency, respectively).

We now turn to the conditional dynamics resulting from the qubit–oscillator coupling. We first determine the blue and red sideband resonant frequencies by spectroscopic means using a two-pulse sequence (Fig. 3a). The qubit is prepared in the excited state by a  $\pi$  pulse at the Larmor frequency. A second pulse (18 ns) of variable frequency induces resonant qubit de-excitation (dips in Fig. 3a, top trace) marking the red sideband and the Larmor frequency. Similarly, after a  $2\pi$  pulse which places the qubit in its ground state, we search for resonant excitations (peaks in Fig. 3a, bottom trace) that mark the Larmor frequency and the blue sideband. No resonance is seen on the red sideband, showing that the oscillator is in its ground state with a large probability. Note that in order to excite the blue sideband, we have to increase the microwave power by at least 10 dB, probably due to less effective microwave transmission in the 8–9 GHz range (note also the absence of spectroscopy peaks in this frequency range in Fig. 1b). At high microwave powers, we observe radiative shifts<sup>20</sup> of the resonances. We now exploit these resonances to study the dynamics of the



**Figure 3** Generation and control of entangled states. **a**, Spectroscopic characterization of the energy levels (see Fig. 1b inset) after a  $\pi$  (upper scan) and a  $2\pi$  (lower scan) Rabi pulse on the qubit transition. In the upper scan, the system is first excited to  $|10\rangle$ , from which it decays towards the  $|01\rangle$  excited state (red sideband at 3.58 GHz) or towards the  $|00\rangle$  ground state ( $F_L = 6.48$  GHz). In the lower scan, the system is rotated back to the initial state  $|00\rangle$ , from which it is excited into the  $|10\rangle$  or  $|11\rangle$  states (see the blue sideband peak (dashed line) at 9.48 GHz for 13 dB more power). **b**, Coupled Rabi oscillations: the blue sideband is excited and the switching probability is recorded as a function of the pulse length for different microwave powers (plots are shifted vertically for clarity). For large microwave powers, the resonance peak of the blue sideband is shifted to 9.15 GHz. When detuning the microwave excitation away from resonance, the Rabi oscillations become faster (bottom four curves). These oscillations are suppressed by preparing the system in the  $|10\rangle$  state with a  $\pi$  pulse and revived after a  $2\pi$  pulse (top two curves in **b**). **c**, Coupled Rabi oscillations: after a  $\pi$  pulse on the qubit resonance ( $|00\rangle \rightarrow |10\rangle$ ) we excite the red sideband at 3.58 GHz. The switching probability shows coherent oscillations between the states  $|10\rangle$  and  $|01\rangle$ , at various microwave powers (the curves are shifted vertically for clarity). The decay time of the coherent oscillations in **a** and **b** is  $\sim 3$  ns.



**Figure 4** Oscillator relaxation time. **a**, Rabi oscillations between the  $|01\rangle$  and  $|10\rangle$  states (during pulse 2 in the inset) obtained after applying a first pulse (1) in resonance with the oscillator transition. Here, the interval between the two pulses is 1 ns. The continuous line represents a fit using an exponentially decaying sinusoidal oscillation plus an exponential decay of the background (due to the relaxation into the ground state). The oscillation's decay time is  $\tau_{\text{coh}} = 2.9$  ns, whereas the background decay time is  $\sim 4$  ns. **b**, The amplitude of Rabi oscillations as a function of the interval between the two pulses (the vertical bars represent standard error bars estimated from the fitting procedure, see **a**). Owing to the oscillator relaxation, the amplitude decays in  $\tau_{\text{rel}} \approx 6$  ns (the continuous line represents an exponential fit).



coupled system by applying pulses of varying length. In Fig. 3b, Rabi oscillations are shown for the  $|00\rangle$  to  $|11\rangle$  transition. When the microwave frequency is detuned from resonance, the Rabi oscillations are accelerated (bottom four curves, to be compared with the fifth curve). After a  $\pi$  pulse which prepares the system in the  $|10\rangle$  state, these oscillations are suppressed (second curve in Fig. 3b). After a  $2\pi$  pulse they are revived (first curve in Fig. 3b). In the case of Fig. 3c, the qubit is first excited onto the  $|10\rangle$  state by a  $\pi$  pulse, and a second pulse in resonance with the red sideband transition drives the system between the  $|10\rangle$  and  $|01\rangle$  states. The Rabi frequency depends linearly on the microwave amplitude, with a smaller slope compared to the bare qubit driving. During the time evolution of the coupled Rabi oscillations shown in Fig. 3b and c, the qubit and the oscillator experience a time-dependent entanglement, although the present data do not permit us to quantify it to a sufficient degree of confidence.

The sideband Rabi oscillations of Fig. 3 show a short coherence time ( $\sim 3$  ns), which we attribute mostly to the oscillator relaxation. To determine its relaxation time, we performed the following experiment. First, we excite the oscillator with a resonant low power microwave pulse. After a variable delay  $\Delta t$ , during which the oscillator relaxes towards  $n = 0$ , we start recording Rabi oscillations on the red sideband transition (see Fig. 4a for  $\Delta t = 1$  ns). The decay of the oscillation amplitude as a function of  $\Delta t$  corresponds to an oscillator relaxation time of  $\sim 6$  ns (Fig. 4b), consistent with a quality factor of 100–150 estimated from the width of the  $\nu_p$  resonance. The exponential fit (continuous line in Fig. 4b) shows an offset of  $\sim 4\%$  due to thermal effects. To estimate the higher bound of the sample temperature, we consider that the visibility of the oscillations presented here (Figs 2–4) is set by the detection efficiency and not by the state preparation. When related to the maximum signal of the qubit Rabi oscillations of  $\sim 40\%$ , the 4%-offset corresponds to  $\sim 10\%$  thermal occupation of oscillator excited states (an effective temperature of  $\sim 60$  mK). Consistently, we also observe low-amplitude red sideband oscillations without preliminary microwave excitation of the oscillator.

We have demonstrated coherent dynamics of a coupled superconducting two-level plus harmonic oscillator system, implying that the two subsystems are entangled. Increasing the coupling strength and the oscillator relaxation time should allow us to quantify the entanglement, as well as to study non-classical states of the oscillator. Our results provide strong indications that solid-state quantum devices could in future be used as elements for the manipulation of quantum information.  $\square$

Received 25 May; accepted 5 July 2004; doi:10.1038/nature02831.

- Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge Univ. Press, Cambridge, 2000).
- Nakamura, Y. *et al.* Coherent control of macroscopic quantum states in a single-Cooper-pair box. *Nature* **398**, 786–788 (1999).
- Vion, D. *et al.* Manipulating the quantum state of an electrical circuit. *Science* **296**, 886–889 (2002).
- Yu, Y., Han, S., Chu, X., Chu, S. & Wang, Z. Coherent temporal oscillations of macroscopic quantum states in a Josephson junction. *Science* **296**, 889–892 (2002).
- Martinis, J. M., Nam, S., Aumentado, J. & Urbina, C. Rabi oscillations in a large Josephson-junction qubit. *Phys. Rev. Lett.* **89**, 117901 (2002).
- Chiorescu, I., Nakamura, Y., Harmans, C. J. P. M. & Mooij, J. E. Coherent quantum dynamics of a superconducting flux qubit. *Science* **299**, 1869–1871 (2003).
- Pashkin, Yu. A. *et al.* Quantum oscillations in two coupled charge qubits. *Nature* **421**, 823–826 (2003).
- Berkley, A. J. *et al.* Entangled macroscopic quantum states in two superconducting qubits. *Science* **300**, 1548–1550 (2003).
- Majer, J. B., Pava, F. G., ter Haar, A. C. J., Harmans, C. J. P. M. & Mooij, J. E. Spectroscopy on two coupled flux qubits. Preprint at (<http://arxiv.org/abs/cond-mat/0308192>) (2003).
- Izmalkov, A. *et al.* Experimental evidence for entangled states formation in a system of two coupled flux qubits. Preprint at (<http://arxiv.org/abs/cond-mat/0312332>) (2003).
- Yamamoto, T., Pashkin, Yu. A., Astafiev, O., Nakamura, Y. & Tsai, J. S. Demonstration of conditional gate operation using superconducting charge qubits. *Nature* **425**, 941–944 (2003).
- Leibfried, D., Blatt, R., Monroe, C. & Wineland, D. Quantum dynamics of single trapped ions. *Rev. Mod. Phys.* **75**, 281–324 (2003).
- Mandel, O. *et al.* Controlled collisions for multi-particle entanglement of optically trapped atoms. *Nature* **425**, 937–940 (2003).
- Raimond, J. M., Brune, M. & Haroche, S. Manipulating quantum entanglement with atoms and photons in a cavity. *Rev. Mod. Phys.* **73**, 565–582 (2001).

- Mooij, J. E. *et al.* Josephson persistent-current qubit. *Science* **285**, 1036–1039 (1999).
- van der Wal, C. H. *et al.* Quantum superposition of macroscopic persistent-current states. *Science* **290**, 773–777 (2000).
- Burkard, G. *et al.* Asymmetry and decoherence in double-layer persistent-current qubit. Preprint at (<http://arxiv.org/abs/cond-mat/0405273>) (2004).
- Goorden, M. C., Thorwart, M. & Grifoni, M. Entanglement spectroscopy of a driven solid-state qubit and its detector. Preprint at (<http://arxiv.org/abs/cond-mat/0405220>) (2004).
- Tinkham, M. *Introduction to Superconductivity* 2nd edn (McGraw-Hill, New York, 1996).
- Cohen-Tannoudji, C., Dupont-Roc, J. & Grynberg, G. *Atom-photon Interactions: Basic Processes and Applications* Ch. II E (Wiley & Sons, New York, 1992).

**Acknowledgements** We thank A. Blais, G. Burkard, D. DiVincenzo, G. Falci, M. Grifoni, S. Lloyd, S. Miyashita, T. Orlando, R. N. Schouten, L. Vandersypen and F. K. Wilhelm for discussions. This work was supported by the Dutch Foundation for Fundamental Research on Matter (FOM), the EU Marie Curie and SQUBIT grants, and the US Army Research Office.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to I.C. (chiorescu@pa.msu.edu) and J.E.M. (mooij@qt.tn.tudelft.nl).

## Strong coupling of a single photon to a superconducting qubit using circuit quantum electrodynamics

A. Wallraff<sup>1</sup>, D. I. Schuster<sup>1</sup>, A. Blais<sup>1</sup>, L. Frunzio<sup>1</sup>, R.-S. Huang<sup>1,2</sup>, J. Majer<sup>1</sup>, S. Kumar<sup>1</sup>, S. M. Girvin<sup>1</sup> & R. J. Schoelkopf<sup>1</sup>

<sup>1</sup>Departments of Applied Physics and Physics, Yale University, New Haven, Connecticut 06520, USA

<sup>2</sup>Department of Physics, Indiana University, Bloomington, Indiana 47405, USA

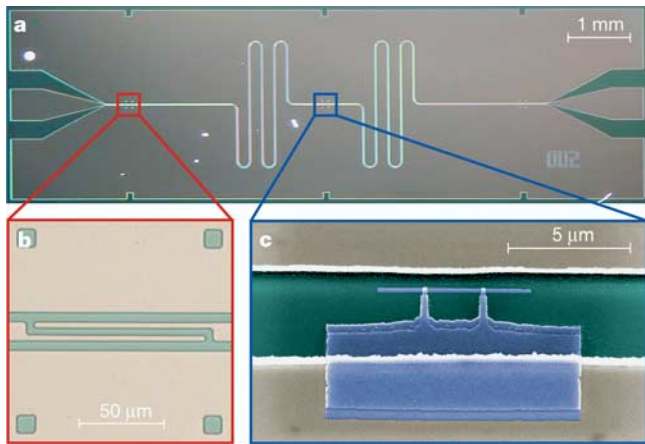
The interaction of matter and light is one of the fundamental processes occurring in nature, and its most elementary form is realized when a single atom interacts with a single photon. Reaching this regime has been a major focus of research in atomic physics and quantum optics<sup>1</sup> for several decades and has generated the field of cavity quantum electrodynamics<sup>2,3</sup>. Here we perform an experiment in which a superconducting two-level system, playing the role of an artificial atom, is coupled to an on-chip cavity consisting of a superconducting transmission line resonator. We show that the strong coupling regime can be attained in a solid-state system, and we experimentally observe the coherent interaction of a superconducting two-level system with a single microwave photon. The concept of circuit quantum electrodynamics opens many new possibilities for studying the strong interaction of light and matter. This system can also be exploited for quantum information processing and quantum communication and may lead to new approaches for single photon generation and detection.

In atomic cavity quantum electrodynamics (QED), an isolated atom with electric dipole moment  $d$  interacts with the vacuum state electric field  $E_0$  of a cavity. The quantum nature of the field gives rise to coherent oscillations of a single excitation between the atom and the cavity at the vacuum Rabi frequency  $\nu_{\text{Rabi}} = 2dE_0/\hbar$ , which can be observed when  $\nu_{\text{Rabi}}$  exceeds the rates of relaxation and decoherence of both the atom and the field. This effect has been observed in the time domain using Rydberg atoms in three-dimensional microwave cavities<sup>3</sup> and spectroscopically using alkali atoms in very small optical cavities with large vacuum fields<sup>4</sup>.

Coherent quantum effects have been recently observed in several superconducting circuits<sup>5–10</sup>, making these systems well suited for use as quantum bits (qubits) for quantum information processing.

Of the various superconducting qubits, the Cooper pair box<sup>11</sup> is especially well suited for cavity QED because of its large effective electric dipole moment  $d$ , which can be  $10^4$  times larger than in an alkali atom and ten times larger than a typical Rydberg atom<sup>12</sup>. As suggested in our earlier theoretical study<sup>12</sup>, the simultaneous combination of this large dipole moment and the large vacuum field strength—due to the small size of the quasi one-dimensional transmission line cavity—in our implementation is ideal for reaching the strong coupling limit of cavity QED in a circuit. Other solid-state analogues of strong coupling cavity QED have been envisaged in superconducting<sup>13–20</sup>, semiconducting<sup>21,22</sup>, and even micro-mechanical systems<sup>23</sup>. First steps towards realizing such a regime have been made for semiconductors<sup>21,24,25</sup>. To our knowledge, our experiments constitute the first experimental observation of strong coupling cavity QED with a single artificial atom and a single photon in a solid-state system.

The on-chip cavity is made by patterning a thin superconducting film deposited on a silicon chip. The quasi-one-dimensional coplanar waveguide resonator<sup>26</sup> consists of a narrow centre conductor of length  $l$  and two nearby lateral ground planes, see Fig. 1a. Close to its full-wave ( $l = \lambda$ ) resonance frequency,  $\omega_r = 2\pi\nu_r = 1/\sqrt{LC} = 2\pi \cdot 6.044$  GHz, where  $\nu_r$  is the bare resonance frequency, the resonator can be modelled as a parallel combination of a capacitor  $C$  and an inductor  $L$  (the internal losses are negligible). This simple resonant circuit behaves as a harmonic oscillator described by the hamiltonian  $H_r = \hbar\omega_r(a^\dagger a + 1/2)$ , where  $\langle a^\dagger a \rangle = \langle \hat{n} \rangle = n$  is the average photon number. At our operating temperature of  $T < 100$  mK, much less than  $\hbar\omega_r/k_B \approx 300$  mK, the resonator is nearly in its ground state, with a thermal occupancy  $n < 0.06$ . The vacuum fluctuations of the resonator give rise to a root mean square (r.m.s.) voltage  $V_{\text{rms}} = \sqrt{\hbar\omega_r/2C} \approx 1 \mu\text{V}$  on its centre conductor,



**Figure 1** Integrated circuit for cavity QED. **a**, The superconducting niobium coplanar waveguide resonator is fabricated on an oxidized  $10 \times 3 \text{ mm}^2$  silicon chip using optical lithography. The width of the centre conductor is  $10 \mu\text{m}$  separated from the lateral ground planes extending to the edges of the chip by a gap of width  $5 \mu\text{m}$  resulting in a wave impedance of the structure of  $Z = 50 \Omega$  being optimally matched to conventional microwave components. The length of the meandering resonator is  $l = 24 \text{ mm}$ . It is coupled by a capacitor at each end of the resonator (see **b**) to an input and output feed line, fanning out to the edge of the chip and keeping the impedance constant. **b**, The capacitive coupling to the input and output lines and hence the coupled quality factor  $Q$  is controlled by adjusting the length and separation of the finger capacitors formed in the centre conductor. **c**, False colour electron micrograph of a Cooper pair box (blue) fabricated onto the silicon substrate (green) into the gap between the centre conductor (top) and the ground plane (bottom) of a resonator (beige) using electron beam lithography and double angle evaporation of aluminium. The Josephson tunnel junctions are formed at the overlap between the long thin island parallel to the centre conductor and the fingers extending from the much larger reservoir coupled to the ground plane.

and an electric field between the centre conductor and the ground plane that is a remarkable  $E_{\text{rms}} \approx 0.2 \text{ V m}^{-1}$ , some hundred times larger than in the three-dimensional cavities used in atomic microwave cavity QED<sup>3</sup>. The large vacuum field strength results from the extremely small effective mode volume ( $\sim 10^{-6}$  cubic wavelengths) of the resonator<sup>12</sup>.

The resonator is coupled via two coupling capacitors  $C_{\text{in/out}}$ , one at each end (see Fig. 1b), to the input and output transmission lines that allow its microwave transmission to be probed (see Fig. 2a–c). The predominant source of dissipation is the loss of photons from the resonator through these ports at a rate  $\kappa = \omega_r/Q$ , where  $Q$  is the (loaded) quality factor of the resonator. The internal (uncoupled) loss of the resonator is negligible ( $Q_{\text{int}} \approx 10^6$ ). Thus, the average photon lifetime in the resonator  $T_r = 1/\kappa$  exceeds 100 ns, even for our initial choice of a moderate quality factor  $Q \approx 10^4$ .

The Cooper pair box (CPB) consists of a several micrometre long and submicrometre wide superconducting island which is coupled via two submicrometre size Josephson tunnel junctions to a much larger superconducting reservoir, and is fabricated in the gap between the centre conductor and the ground plane of the resonator, at an antinode of the field (see Fig. 1c). The CPB is a two-state system described by the hamiltonian<sup>13</sup>  $H_a = -(E_{\text{el}}\sigma_x + E_J\sigma_z)/2$ , where  $E_{\text{el}} = 4E_C(1 - n_g)$  is the electrostatic energy and  $E_J = E_{J,\text{max}}\cos(\pi\Phi_b)$  is the Josephson energy. The overall energy scales of these terms, the charging energy  $E_C$  and the Josephson energy  $E_{J,\text{max}}$ , can be readily engineered during the fabrication by the choice of the total box capacitance and resistance respectively, and then further tuned in situ by electrical means. A gate voltage  $V_g$  applied to the input port (see Fig. 2a), induces a gate charge  $n_g = V_g C_g^*/e$  that controls  $E_{\text{el}}$ , where  $C_g^*$  is the effective capacitance between the input port of the resonator and the island of the CPB. A flux bias  $\Phi_b = \Phi/\Phi_0$ , applied with an external coil to the loop of the box, controls  $E_J$ . Denoting the ground state of the box as  $|\downarrow\rangle$  and the first excited state as  $|\uparrow\rangle$  (see Fig. 2d), we have a two-level system whose energy separation  $E_a = \hbar\omega_a$  can be widely varied as shown in Fig. 3c. Coherence of the CPB is limited by relaxation from the excited state at a rate  $\gamma_1$ , and by fluctuations of the level separation giving rise to dephasing at a rate  $\gamma_\varphi$ , for a total decoherence rate  $\gamma = \gamma_1/2 + \gamma_\varphi$  (ref. 13).

The Cooper pair box couples to photons stored in the resonator by an electric dipole interaction, via the coupling capacitance  $C_g$ . The vacuum voltage fluctuations  $V_{\text{rms}}$  on the centre conductor of the resonator change the energy of a Cooper pair on the box island by an amount  $\hbar g = dE_0 = eV_{\text{rms}}C_g/C_\Sigma$ . We have shown<sup>12</sup> that this coupled system is described by the Jaynes–Cummings hamiltonian  $H_{\text{JC}} = H_r + H_a + \hbar g(a^\dagger\sigma^- + a\sigma^+)$ , where  $\sigma^+$  ( $\sigma^-$ ) creates (annihilates) an excitation in the CPB. It describes the coherent exchange of energy between a quantized electromagnetic field and a quantum two-level system at a rate  $g/2\pi$ , which is observable if  $g$  is much larger than the decoherence rates  $\gamma$  and  $\kappa$ . This strong coupling limit<sup>3</sup>  $g > [\gamma, \kappa]$  is achieved in our experiments. When the detuning  $\Delta = \omega_a - \omega_r$  is equal to zero, the eigenstates of the coupled system are symmetric and antisymmetric superpositions of a single photon and an excitation in the CPB  $|\pm\rangle = (|0,1\rangle \pm |1,0\rangle)/\sqrt{2}$  with energies  $E_\pm = \hbar(\omega_r \pm g)$ . Although the cavity and the CPB are entangled in the eigenstates  $|\pm\rangle$ , their entangled character is not addressed in our current cavity QED experiment which spectroscopically probes the energies  $E_\pm$  of the coherently coupled system.

The strong coupling between the field in the resonator and the CPB can be used to perform a quantum nondemolition (QND) measurement of the state of the CPB in the non-resonant (dispersive) limit  $|\Delta| \gg g$ . Diagonalization of the coupled quantum system leads to the effective hamiltonian<sup>12</sup>:

$$H \approx \hbar \left( \omega_r + \frac{g^2}{\Delta} \right) a^\dagger a + \frac{1}{2} \hbar \left( \omega_a + \frac{g^2}{\Delta} \right) \sigma_z$$

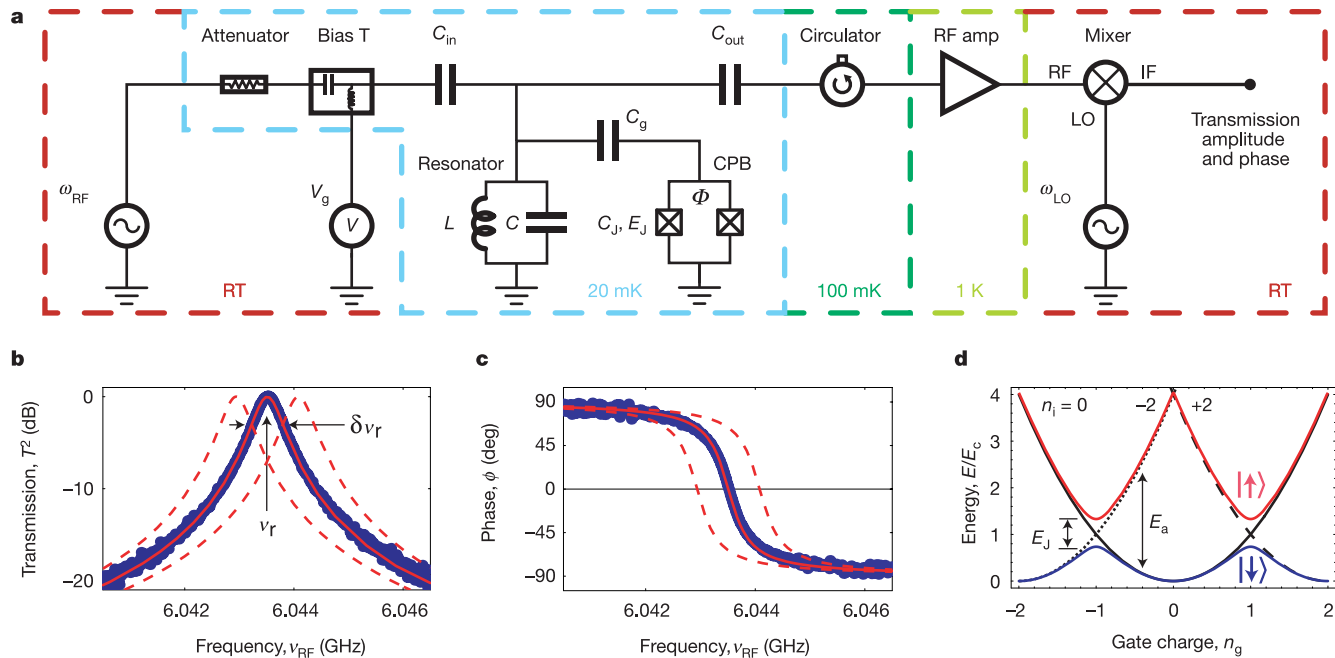
The transition frequency  $\omega_r \pm g^2/\Delta$  is now conditioned by the qubit state  $\sigma_z = \pm 1$ . Thus, by measuring the transition frequency of the resonator, the qubit state can be determined. Similarly, the level separation in the qubit  $\hbar(\omega_a + 2a^\dagger a g^2/\Delta + g^2/\Delta)$  depends on the number of photons in the resonator. The term  $2a^\dagger a g^2/\Delta$ , linear in  $\hat{n}$ , is the alternating current (a.c.) Stark shift and  $g^2/\Delta$  is the Lamb shift. All terms in this hamiltonian, with the exception of the Lamb shift, are clearly identified in the results of our circuit QED experiments.

The properties of this coupled system are determined by probing the resonator spectroscopically<sup>12</sup>. The amplitude  $T$  and phase  $\phi$  of a microwave probe beam of power  $P_{\text{RF}}$  transmitted through the resonator are measured versus probe frequency  $\omega_{\text{RF}}$ . A simplified schematic of the microwave circuit is shown in Fig. 2a. In this setup, the CPB acts as an effective capacitance that is dependent on its  $\sigma_z$  eigenstate, the coupling strength  $g$ , and detuning  $\Delta$ . This variable capacitance changes the resonator frequency and its transmission spectrum. The transmission  $T^2$  and phase  $\phi$  of the resonator for a far-detuned qubit ( $g^2/\kappa\Delta \ll 1$ ), that is, when the qubit is effectively decoupled from the resonator, are shown in Fig. 2b and c. In this case, the transmission is a lorentzian of width  $\delta\nu_r = \nu_r/Q = \kappa/2\pi$  at  $\nu_r$ , and the phase  $\phi$  displays a corresponding step of  $\pi$ . The expected transmission at smaller detuning corresponding to a frequency shift  $\pm g^2/\Delta = \kappa$  are shown by dashed lines in Fig. 2b and c. Such small shifts in the resonator frequency are sensitively measured as a phase shift  $\phi = \pm \tan^{-1}(2g^2/\kappa\Delta)$  of the transmitted microwave at a fixed

probe frequency  $\omega_{\text{RF}}$  using beam powers  $P_{\text{RF}}$  which controllably populate the resonator with average photon numbers from  $n \approx 10^3$  down to the sub-photon level  $n \ll 1$ . We note that both the resonator and qubit can be controlled and measured using capacitive and inductive coupling only, that is, without attaching any d.c. connections to either system.

Measurements of the phase  $\phi$  versus  $n_g$  are shown in Fig. 3b, and two different cases can be identified for a Cooper pair box with Josephson energy  $E_{J,\text{max}}/\hbar > \nu_r$ . In the first case, for bias fluxes such that  $E_J(\Phi_b)/\hbar > \nu_r$ , the qubit does not come into resonance with the resonator for any value of gate charge  $n_g$  (see Fig. 3a). As a result, the measured phase shift  $\phi$  is maximum for the smallest detuning  $\Delta$  at  $n_g = 1$  and gets smaller as  $\Delta$  increases (see Fig. 3b). Moreover,  $\phi$  is periodic in  $n_g$  with a period of  $2e$ , as expected. In the second case, for values of  $\Phi_b$  resulting in  $E_J(\Phi_b)/\hbar < \nu_r$ , the qubit goes through resonance with the resonator at two values of  $n_g$ . Thus, the phase shift  $\phi$  is largest as the qubit approaches resonance ( $\Delta \rightarrow 0$ ) at the points indicated by red arrows (see Fig. 3a, b). As the qubit goes through resonance, the phase shift  $\phi$  changes sign when  $\Delta$  changes sign. This behaviour is in perfect agreement with predictions based on the analysis of the circuit QED hamiltonian in the dispersive regime.

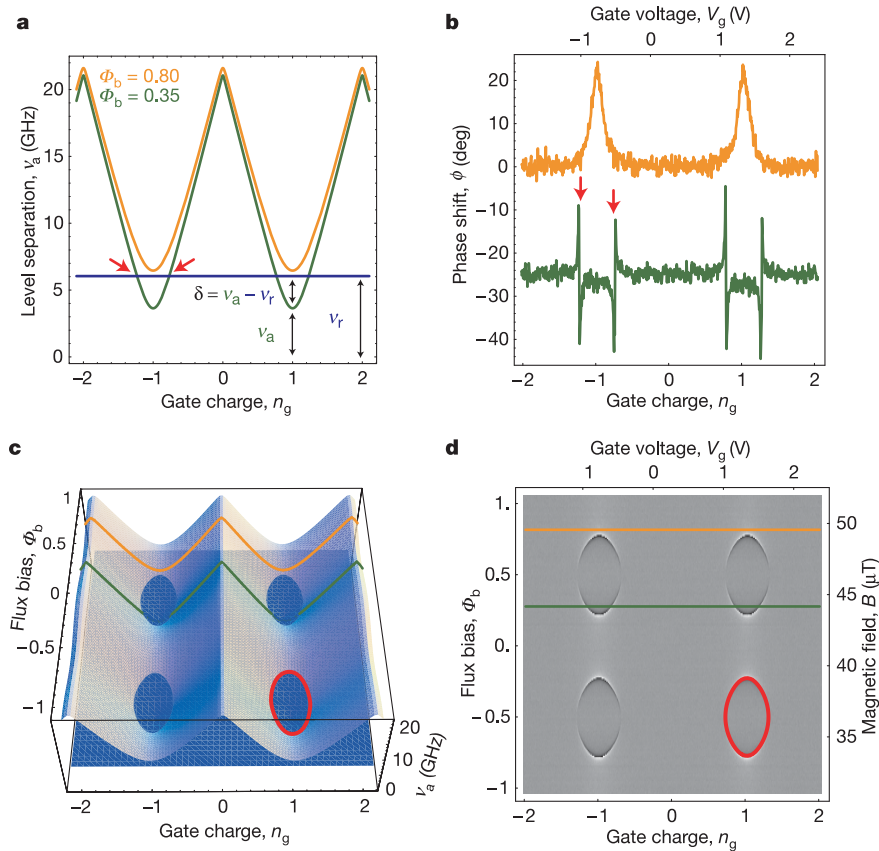
In Fig. 3c the qubit level separation  $\nu_a = E_a/\hbar$  is plotted versus the bias parameters  $n_g$  and  $\Phi_b$ . The qubit is in resonance with the resonator at the points  $[n_g, \Phi_b]$ , indicated by the red curve in one quadrant of the plot. The measured phase shift  $\phi$  is plotted versus



**Figure 2** Measurement scheme, resonator and Cooper pair box. **a**, The resonator with effective inductance  $L$  and capacitance  $C$  coupled through the capacitor  $C_0$  to the Cooper pair box with junction capacitance  $C_J$  and Josephson energy  $E_J$  forms the circuit QED system which is coupled through  $C_{\text{in/out}}$  to the input/output ports. The value of  $E_J$  is controllable by the magnetic flux  $\Phi$ . The input microwave at frequency  $\omega_{\text{RF}}$  is added to the gate voltage  $V_g$  using a bias-tee. After the transmitted signal at  $\omega_{\text{RF}}$  is amplified using a cryogenic high electron mobility (HEMT) amplifier and mixed with the local oscillator at  $\omega_{\text{LO}}$ , its amplitude and phase are determined. The circulator and the attenuator prevent leakage of thermal radiation into the resonator. The temperature of individual components is indicated. **b**, Measured transmission power spectrum of the resonator (blue dots), the full linewidth  $\delta\nu_r$  at half-maximum and the centre frequency  $\nu_r$  are indicated. The solid red line is a fit to a lorentzian with  $Q = \nu_r/\delta\nu_r \approx 10^4$ . **c**, Measured transmission phase  $\phi$  (blue dots) with fit (red line). In panels **b** and **c** the dashed lines are theory curves shifted by

$\pm \delta\nu_r$  with respect to the data. **d**, Energy level diagram of a Cooper pair box. The electrostatic energy  $E_c(n_i - n_g)^2$ , with charging energy  $E_c = e^2/2C_\Sigma$ , is indicated for  $n_i = 0$  (solid black line),  $-2$  (dotted line) and  $+2$  (dashed line) excess electrons forming Cooper pairs on the island.  $C_\Sigma$  is the total capacitance of the island given by the sum of the capacitances  $C_J$  of the two tunnel junctions, the coupling capacitance  $C_0$  to the centre conductor of the resonator and any stray capacitances. In the absence of Josephson tunnelling the states with  $n_i$  and  $n_i + 2$  electrons on the island are degenerate at  $n_g = 1$ . The Josephson coupling mediated by the weak link formed by the tunnel junctions between the superconducting island and the reservoir lifts this degeneracy and opens up a gap proportional to the Josephson energy  $E_J = E_{J,\text{max}} \cos(\pi\Phi_b)$ , where  $E_{J,\text{max}} = \hbar\Delta_A/8e^2R_J$ , with the superconducting gap of aluminium  $\Delta_A$  and the tunnel junction resistance  $R_J$ . A ground-state band ( $\downarrow\downarrow$ ) and an excited-state band ( $\uparrow\uparrow$ ) are formed with a gate charge and flux-bias-dependent energy level separation of  $E_a$ .





**Figure 3** Strong coupling circuit QED in the dispersive regime. **a**, Calculated level separation  $\nu_a = \omega_a/2\pi = E_a/h$  between ground  $|\downarrow\rangle$  and excited state  $|\uparrow\rangle$  of qubit for two values of flux bias  $\Phi_b = 0.8$  (orange line) and  $\Phi_b = 0.35$  (green line). The resonator frequency  $\nu_r = \omega_r/2\pi$  is shown by a blue line. Resonance occurs at  $\nu_a = \nu_r$  symmetrically around degeneracy  $n_g = \pm 1$ ; also see red arrows. The detuning  $\Delta/2\pi = \delta = \nu_a - \nu_r$  is indicated. **b**, Measured phase shift  $\phi$  of the transmitted microwave for values of  $\Phi_b$  in **a**. Green curve is offset by  $-25$  deg for visibility. **c**, Calculated qubit level separation  $\nu_a$  versus bias parameters  $n_g$  and  $\Phi_b$ . The resonator frequency  $\nu_r$  is indicated by the blue plane. At the intersection, also indicated by the red

curve in the lower right-hand quadrant, resonance between the qubit and the resonator occurs ( $\delta = 0$ ). For qubit states below the resonator plane the detuning is  $\delta < 0$ , above  $\delta > 0$ . **d**, Density plot of measured phase shift  $\phi$  versus  $n_g$  and  $\Phi_b$ . Light colours indicate positive  $\phi$  ( $\delta > 0$ ), dark colours negative  $\phi$  ( $\delta < 0$ ). The red line is a fit of the data to the resonance condition  $\nu_a = \nu_r$ . In **c** and **d**, the line cuts presented in **a** and **b** are indicated by the orange and the green line, respectively. The microwave probe power  $P_{RF}$  used to acquire the data is adjusted such that the maximum intra-resonator photon number  $n$  at  $\nu_r$  is about ten for  $g^2/\kappa\Delta \ll 1$ . The calibration of the photon number has been performed in situ by measuring the a.c.-Stark shift of the qubit levels.

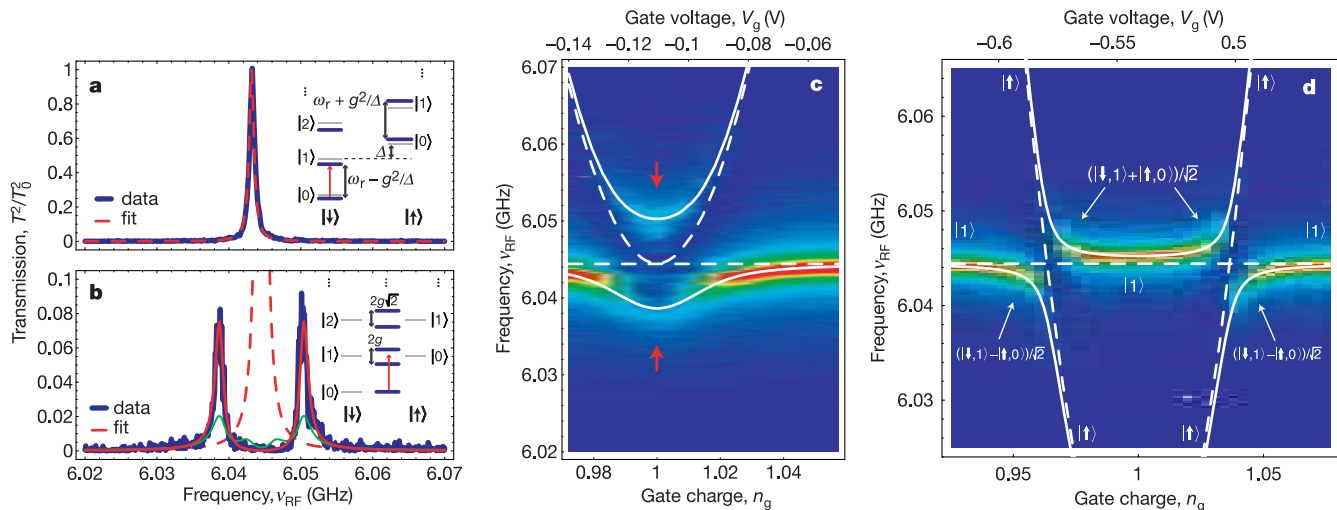
both  $n_g$  and  $\Phi_b$  in Fig. 3d. We observe the expected periodicity in flux bias  $\Phi_b$  with one flux quantum  $\Phi_0$ . The set of parameters  $[n_g, \Phi_b]$  for which the resonance condition is met is marked by a sudden sign change in  $\phi$ , which allows a determination of the Josephson energy  $E_{J,\max} = 8.0 (\pm 0.1)$  GHz and the charging energy  $E_C = 5.2 (\pm 0.1)$  GHz.

These data clearly demonstrate that the properties of the qubit can be determined in a transmission measurement of the resonator and that full in situ control over the qubit parameters is achieved. We note that in the dispersive regime this new read-out scheme for the Cooper pair box is most sensitive at charge degeneracy ( $n_g = 1$ ), where the qubit is to first order decoupled from  $1/f$  fluctuations in its charge environment, which minimizes dephasing<sup>6</sup>. This property is advantageous for quantum control of the qubit at  $n_g = 1$ , a point where traditional electrometry, using a single electron transistor (SET) for example<sup>27</sup>, is unable to distinguish the qubit states. We note that this dispersive QND measurement of the qubit state<sup>12</sup> is the complement of the atomic microwave cavity QED measurement in which the state of the cavity is inferred non-destructively from the phase shift in the state of a beam of atoms sent through the cavity<sup>3,28</sup>.

Making use of the full control over the qubit hamiltonian, we then tune the flux bias  $\Phi_b$  so that the qubit is at  $n_g = 1$  and in resonance with the resonator. Initially, the resonator and the qubit are cooled into their combined ground state  $|0, \downarrow\rangle$ ; see inset in

Fig. 4b. Owing to the coupling, the first excited states become a doublet  $|\pm\rangle$ . Similarly to ref. 4, we probe the energy splitting of this doublet spectroscopically using a weak probe beam so that  $n \ll 1$ . The intra-resonator photon number,  $n$ , is calibrated by measuring the a.c.-Stark shift of the qubit in the dispersive case. The resonator transmission  $T^2$  is first measured for large detuning  $\Delta$  with a probe beam populating the resonator with a maximum of  $n \approx 1$  at resonance; see Fig. 4a. From the lorentzian line the photon decay rate of the resonator is determined as  $\kappa/2\pi = 0.8$  MHz. The probe beam power is subsequently reduced by 5 dB and the transmission spectrum  $T^2$  is measured in resonance ( $\Delta = 0$ ); see Fig. 4b. We clearly observe two well-resolved spectral lines separated by the vacuum Rabi frequency  $\nu_{\text{Rabi}} \approx 11.6$  MHz. The individual lines have a width determined by the average of the photon decay rate  $\kappa$  and the qubit decoherence rate  $\gamma$ . The data are in excellent agreement with the transmission spectrum numerically calculated using the given value  $\kappa/2\pi = 0.8$  MHz and the single adjustable parameter  $\gamma/2\pi = 0.7$  MHz.

The transmission spectrum shown in Fig. 4b is highly sensitive to the photon number in the cavity. The measured transmission spectrum is consistent with the expected thermal photon number of  $n \lesssim 0.06$  ( $T < 100$  mK); see red curve in Fig. 4b. Owing to the anharmonicity of the coupled atom-cavity system in the resonant case, an increased thermal photon number would reduce trans-



**Figure 4** Vacuum Rabi mode splitting. **a**, Measured transmission  $T^2$  (blue line) versus microwave probe frequency  $\nu_{\text{RF}}$  for large detuning ( $g^2/\Delta \ll 1$ ) and fit to Lorentzian (dashed red line). The peak transmission amplitude is normalized to unity. The inset shows the dispersive dressed states level diagram. **b**, Measured transmission spectrum for the resonant case  $\Delta = 0$  at  $n_g = 1$  (blue line) showing the vacuum Rabi mode splitting compared to numerically calculated transmission spectra (red and green lines) for thermal photon numbers of  $n = 0.06$  and  $0.5$ , respectively. The dashed red line is the calculated transmission for  $g = 0$  and  $\kappa/2\pi = 0.8$  MHz. The inset shows the resonant dressed

states level diagram. **c**, Resonator transmission amplitude  $T$  plotted versus probe frequency  $\nu_{\text{RF}}$  and gate charge  $n_g$  for  $\Delta = 0$  at  $n_g = 1$ . Blue colour corresponds to small  $T$ , red colour to large  $T$ . Dashed lines are uncoupled qubit level separation  $\nu_a$  and resonator resonance frequency  $\nu_r$ . Solid lines are level separations found from exact diagonalization of  $H_{\text{JC}}$ . Spectrum shown in **b** corresponds to line cut along red arrows. **d**, As in **c**, but for  $E_J/h < \nu_r$ . The dominant character of the corresponding eigenstates is indicated.

mission and give rise to additional peaks in the spectrum owing to transitions between higher excited doublets<sup>30</sup>. The transmission spectrum calculated for a thermal photon number of  $n = 0.5$  (see green curve in Fig. 4b) is clearly incompatible with our experimental data, indicating that the coupled system has in fact cooled to near its ground state, and that we measure the coupling of a single qubit to a single photon. The nonlinearity of the cavity QED system is also observed at higher probe beam powers, as transitions are driven between states higher up the dressed state ladders (not shown).

We also observe the anti-crossing between the single photon resonator state and the first excited qubit state by tuning the qubit into and out of resonance with a gate charge near  $n_g = 1$  and measuring the transmission spectrum (see Fig. 4c). The vacuum Rabi peaks evolve from a state with equal weight in the photon and qubit at  $n_g = 1$  (as shown in Fig. 4b) to predominantly photon states for  $n_g \gg 1$  or  $n_g \ll 1$ . The observed peak positions agree well with calculations considering the qubit with level separation  $\nu_a$ , a single photon in the resonator with frequency  $\nu_r$  and a coupling strength of  $g/2\pi$ ; see solid lines in Fig. 4c. For a different value of flux bias  $\Phi_b$  such that  $E_a/h < \nu_r$  at  $n_g = 1$ , two anti-crossings are observed (see Fig. 4d) again in agreement with theory.

The observation of the vacuum Rabi mode splitting and the corresponding avoided crossings demonstrates that the strong coupling limit of cavity QED has been achieved, and that coherent superpositions of a single qubit and a single photon can be generated on a superconducting chip. This opens up many new possibilities for quantum optical experiments with circuits. Possible applications include using the cavity as a quantum bus to couple widely separated qubits in a quantum computer, or as a quantum memory to store quantum information, or even as a generator and detector of single microwave photons for quantum communication.  $\square$

Received 11 June; accepted 12 July 2004; doi:10.1038/nature02851.

- Walls, D. & Milburn, G. *Quantum Optics* (Springer, Berlin, 1994).
- Mabuchi, H. & Doherty, A. Cavity quantum electrodynamics: Coherence in context. *Science* **298**, 1372–1377 (2002).
- Raimond, J., Brune, M. & Haroche, S. Manipulating quantum entanglement with atoms and photons in a cavity. *Rev. Mod. Phys.* **73**, 565–582 (2001).

- Thompson, R. J., Rempe, G. & Kimble, H. J. Observation of normal-mode splitting for an atom in an optical cavity. *Phys. Rev. Lett.* **68**, 1132–1135 (1992).
- Nakamura, Y., Pashkin, Y. A. & Tsai, J. S. Coherent control of macroscopic quantum states in a single-Cooper-pair box. *Nature* **398**, 786–788 (1999).
- Vion, D. *et al.* Manipulating the quantum state of an electrical circuit. *Science* **296**, 886–889 (2002).
- Martinis, J. M., Nam, S., Aumentado, J. & Urbina, C. Rabi oscillations in a large Josephson-junction qubit. *Phys. Rev. Lett.* **89**, 117901 (2002).
- Chiorescu, I., Nakamura, Y., Harmans, C. J. P. M. & Mooij, J. E. Coherent quantum dynamics of a superconducting flux qubit. *Science* **299**, 1869–1871 (2003).
- Yamamoto, T., Pashkin, Y. A., Astafiev, O., Nakamura, Y. & Tsai, J. S. Demonstration of conditional gate operation using superconducting charge qubits. *Nature* **425**, 941–944 (2003).
- Berkley, A. J. *et al.* Entangled macroscopic quantum states in two superconducting qubits. *Science* **300**, 1548–1550 (2003).
- Bouchiat, V., Vion, D., Joyez, P., Esteve, D. & Devoret, M. H. Quantum coherence with a single Cooper pair. *Phys. Scr.* **T76**, 165–170 (1998).
- Blais, A., Huang, R.-S., Wallraff, A., Girvin, S. & Schoelkopf, R. Cavity quantum electrodynamics for superconducting electrical circuits: an architecture for quantum computation. *Phys. Rev. A* **69**, 062320 (2004).
- Makhlin, Y., Schön, G. & Shnirman, A. Quantum-state engineering with Josephson-junction devices. *Rev. Mod. Phys.* **73**, 357–400 (2001).
- Buisson, O. & Hekking, F. in *Macroscopic Quantum Coherence and Quantum Computing* (eds Averin, D. V., Ruggiero, B. & Silvestrini, P.) (Kluwer, New York, 2001).
- Marquardt, F. & Bruder, C. Superposition of two mesoscopically distinct quantum states: Coupling a Cooper-pair box to a large superconducting island. *Phys. Rev. B* **63**, 054514 (2001).
- Al-Saidi, W. A. & Stroud, D. Eigenstates of a small Josephson junction coupled to a resonant cavity. *Phys. Rev. B* **65**, 014512 (2001).
- Plastina, F. & Falcì, G. Communicating Josephson qubits. *Phys. Rev. B* **67**, 224514 (2003).
- Blais, A., Maassen van den Brink, A. & Zagoskin, A. Tunable coupling of superconducting qubits. *Phys. Rev. Lett.* **90**, 127901 (2003).
- Yang, C.-P., Chu, S.-I. & Han, S. Possible realization of entanglement, logical gates, and quantum-information transfer with superconducting-quantum-interference-device qubits in cavity QED. *Phys. Rev. A* **67**, 042311 (2003).
- You, J. Q. & Nori, F. Quantum information processing with superconducting qubits in a microwave field. *Phys. Rev. B* **68**, 064509 (2003).
- Kiraz, A. *et al.* Cavity-quantum electrodynamics using a single InAs quantum dot in a microdisk structure. *Appl. Phys. Lett.* **78**, 3932–3934 (2001).
- Childress, L., Sørensen, A. S. & Lukin, M. D. Mesoscopic cavity quantum electrodynamics with quantum dots. *Phys. Rev. A* **69**, 042302 (2004).
- Irish, E. K. & Schwab, K. Quantum measurement of a coupled nanomechanical resonator–Cooper-pair box system. *Phys. Rev. B* **68**, 155311 (2003).
- Weisbuch, C., Nishioka, M., Ishikawa, A. & Arakawa, Y. Observation of the coupled exciton-photon mode splitting in a semiconductor quantum microcavity. *Phys. Rev. Lett.* **69**, 3314–3317 (1992).
- Vuckovic, J., Fattal, D., Santori, C., Solomon, G. S. & Yamamoto, Y. Enhanced single-photon emission from a quantum dot in a micropost microcavity. *Appl. Phys. Lett.* **82**, 3596 (2003).
- Day, P. K., LeDuc, H. G., Mazin, B. A., Vayonakis, A. & Zmuidzinas, J. A broadband superconducting detector suitable for use in large arrays. *Nature* **425**, 817–821 (2003).
- Lehnert, K. *et al.* Measurement of the excited-state lifetime of a microelectronic circuit. *Phys. Rev. Lett.* **90**, 027002 (2003).

28. Nogues, G. *et al.* Seeing a single photon without destroying it. *Nature* **400**, 239–242 (1999).
29. Schuster, D. I. *et al.* AC-Stark shift and dephasing of a superconducting qubit strongly coupled to a cavity field. Preprint at <http://www.arXiv.org/cond-mat/0408367> (2004).
30. Rau, I., Johansson, G. & Shnirman, A. Cavity QED in superconducting circuits: susceptibility at elevated temperatures. Preprint at <http://www.arXiv.org/cond-mat/0403257> (2004).

**Acknowledgements** We thank J. Teufel, B. Turek and J. Wyatt for their contributions to the project and are grateful to P. Day, D. DeMille, M. Devoret, S. Weinreb and J. Zmuidzinas for numerous conversations. This work was supported in part by the National Security Agency and Advanced Research and Development Activity under the Army Research Office, the NSF, the David and Lucile Packard Foundation, the W. M. Keck Foundation, and the Natural Science and Engineering Research Council of Canada.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to A. W. ([andreas.wallraff@yale.edu](mailto:andreas.wallraff@yale.edu)).

## Generation of ultraviolet entangled photons in a semiconductor

Keiichi Edamatsu<sup>1,2</sup>, Goro Oohata<sup>1,3</sup>, Ryosuke Shimizu<sup>2</sup> & Tadashi Itoh<sup>4,2</sup>

<sup>1</sup>Research Institute of Electrical Communication, Tohoku University, Sendai 980-8577, Japan

<sup>2</sup>CREST, Japan Science and Technology Agency (JST), Japan

<sup>3</sup>ERATO Semiconductor Spintronics Project, JST, Japan

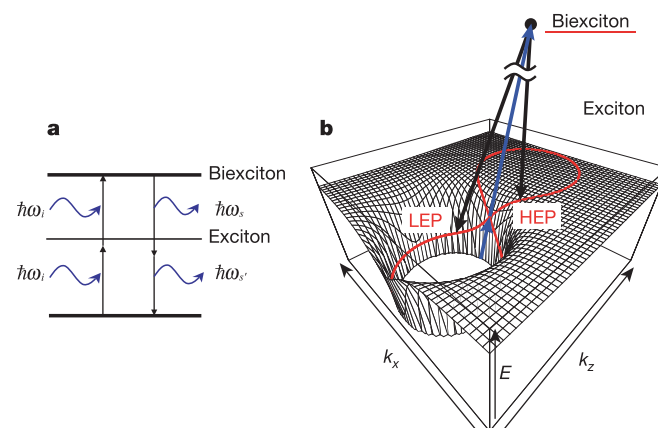
<sup>4</sup>Graduate School of Engineering Science, Osaka University, Toyonaka 560-8531, Japan

Entanglement is one of the key features of quantum information and communications technology. The method that has been used most frequently to generate highly entangled pairs of photons<sup>1,2</sup> is parametric down-conversion. Short-wavelength entangled photons are desirable for generating further entanglement between three or four photons, but it is difficult to use parametric down-conversion to generate suitably energetic entangled photon pairs. One method that is expected to be applicable for the generation of such photons<sup>3</sup> is resonant hyper-parametric scattering (RHPS): a pair of entangled photons is generated in a semiconductor via an electronically resonant third-order nonlinear optical process. Semiconductor-based sources of entangled photons would also be advantageous for practical quantum technologies, but attempts to generate entangled photons in semiconductors have not yet been successful<sup>4,5</sup>. Here we report experimental evidence for the generation of ultraviolet entangled photon pairs by means of biexciton resonant RHPS in a single crystal of the semiconductor CuCl. We anticipate that our results will open the way to the generation of entangled photons by current injection, analogous to current-driven single photon sources<sup>6,7</sup>.

The material we used in this study was copper chloride (CuCl) single crystal. Because CuCl has a large bandgap ( $\sim 3.4$  eV), it is suitable for generating photon pairs in the short wavelength region near ultraviolet. Furthermore, the material has large binding energies for the exciton ( $\sim 200$  meV) and biexciton ( $\sim 30$  meV). These characteristics have made CuCl one of the most thoroughly investigated materials on the physics of excitons and biexcitons (ref. 8 and references therein). In particular, the ‘giant oscillator strength’ in the two-photon excitation of the biexciton results in a large increase in RHPS efficiency, which is advantageous for our experiment. In fact the RHPS in CuCl has been observed since the 1970s (refs 8, 9 and ref. 10 and references therein). Figure 1a schematically shows the RHPS process in resonance to the biexciton state. The two pump (parent) photons (frequency  $\omega_i$ ) resonantly create the

biexciton, and are converted into the two scattered (daughter) photons ( $\omega_s, \omega_s'$ ). The biexciton state ( $\Gamma_1$ ) created in this process has zero angular momentum ( $J = 0$ ), so we expected the polarizations of the daughter photons to be entangled so that their total angular momentum is also zero. With this expectation in mind, we note that polarization correlation between two classical pump beams has been known since the early 1980s (ref. 11). In practice, instead of the oversimplified picture in Fig. 1a, we must consider the exciton-polariton picture; the RHPS obeys the phase-matching condition that takes into account the polariton dispersion relation<sup>8</sup>. The RHPS in this case is also called two-photon resonant polariton scattering or spontaneous hyper-Raman scattering. In this process, shown in Fig. 1b, the biexciton is created from a pair of parent photons (polaritons, more accurately). The sum of the parent photons’ energies matches the biexciton energy. The biexciton progressively coherently decays into two polaritons, the sum of whose photon energies, as well as the sum of momenta, is conserved as that of the biexciton. Although the RHPS in CuCl has been known for decades, the possibility of generating entangled photons by this process was theoretically pointed out only lately<sup>12</sup>. In addition, a large parametric gain via the biexcitonic resonance in CuCl was reported recently<sup>13</sup>. Similar stimulated parametric scattering of polaritons has also been observed in semiconductor microcavities, even at high temperatures<sup>14</sup>.

In the present experiment, we used a vapour-phase-grown thin single crystal of CuCl. Figure 2 presents the schematic drawing of our experimental set-up and Fig. 3 shows the spectrum of light emitted from the sample. The large peak at the downward arrow in Fig. 3 is the Rayleigh scattered light of the pump beam that was tuned to the two-photon excitation resonance of the biexciton. The two peaks indicated by LEP and HEP (lower and higher energy polaritons) on either side of the pump beam originate from the RHPS. The RHPS is very efficient (a few orders of magnitude higher than that of typical parametric down-conversion): We got of the order of  $10^{10}$  photons  $s^{-1} sr^{-1}$  by using pump light of  $\sim 2$  mW. A pair of photons, one from LEP and the other from HEP, is emitted into different directions according to the phase-matching condition, so we placed two optical fibres at appropriate positions and led each photon within the pair into two independent monochromators followed by two photomultipliers (PMTs). A time-interval analyser recorded the time interval ( $\tau$ ) between the detected



**Figure 1** Schematic diagram of the resonant hyper-parametric scattering (RHPS) via biexciton. **a**, Two pump (parent) photons of frequency  $\omega_i$  are converted to the two scattered (daughter) photons ( $\omega_s, \omega_{s'}$ ). **b**, The polariton dispersion drawn in two dimensions of momentum space. The biexciton decays into two polaritons that satisfy the phase-matching condition so that both energy and momentum are conserved. The red curve on the polariton-dispersion surface indicates the states on which the phase-matching condition can be satisfied.

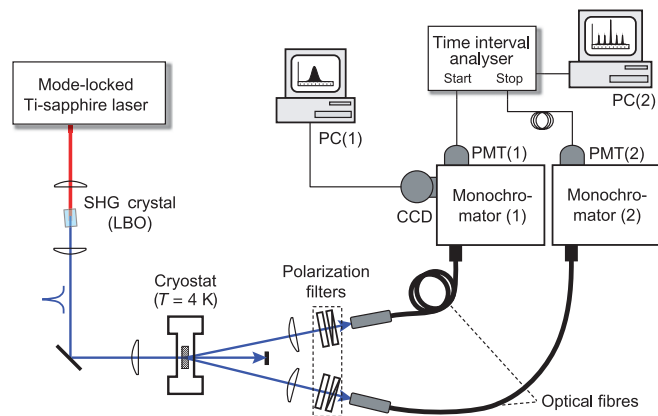


photons. As shown in Fig. 4, we confirmed that the signal was significantly enhanced at  $\tau = 0$ , indicating that we detected the time-correlated photon pairs generated from one biexciton. The other peaks at  $\tau \neq 0$  are caused by uncorrelated pairs of photons emitted from independent biexcitons excited by the train of pump pulses. We expected the main peak at  $\tau = 0$  also to contain practically the same amount of uncorrelated component originating from two biexcitons excited by a single pump pulse. Thus, in the analyses described below, the uncorrelated component has been subtracted from the measured signal.

As mentioned above, the photon pair is expected to have polarization entanglement such that the two-photon polarization state  $|\Psi\rangle$  is expressed as:

$$\begin{aligned} |\Psi\rangle &= \frac{1}{\sqrt{2}}(|L_1 R_2\rangle + |R_1 L_2\rangle) \\ &= \frac{1}{\sqrt{2}}(|H_1 H_2\rangle + |V_1 V_2\rangle) \\ &= \frac{1}{\sqrt{2}}(|D_1 D_2\rangle + |\bar{D}_1 \bar{D}_2\rangle) \end{aligned} \quad (1)$$

where  $L_i$ ,  $R_i$ ,  $H_i$ ,  $V_i$ ,  $D_i$  and  $\bar{D}_i$  represents the states for left-circular, right-circular, horizontal, vertical,  $+45^\circ$  and  $-45^\circ$  polarizations of one of the daughter photons (indicated by  $i = 1, 2$ ), respectively. Hereafter, we omit the indices  $i$  for simplicity. For the first check of the entanglement of equation (1), we measured the time-correlated coincidence traces for the combinations of linear polarized detections, as shown in Fig. 5. The result showed that the coincidences at  $\tau = 0$  were significantly enhanced for  $HH$ ,  $VV$ ,  $DD$  and  $\bar{D}\bar{D}$  polarizations, but not for  $HV$ ,  $VH$ ,  $D\bar{D}$  or  $\bar{D}D$  in accordance with (1). Furthermore, by measuring the signal for  $\geq 16$  combinations of detection polarizations in  $H$ ,  $V$ ,  $L$ ,  $R$ ,  $D$ , and  $\bar{D}$ , we performed maximum-likelihood quantum state tomography<sup>15</sup> to obtain the density matrix  $\hat{\rho}$  of the two-photon polarization state. Figure 6 presents the real and imaginary parts of  $\hat{\rho}$  thus reconstructed. We



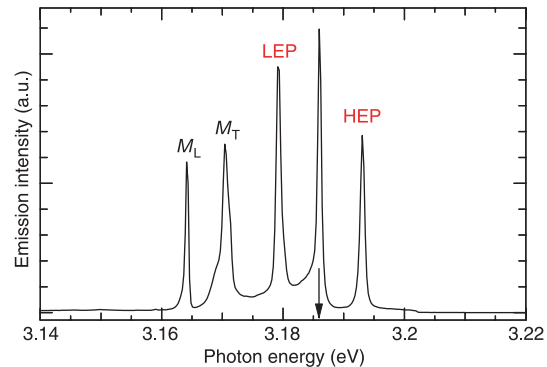
**Figure 2** Experimental set-up to measure the photon correlation of the RHPS. A sample (CuCl thin crystal) was mounted in the cryostat, and its temperature was kept at  $\sim 4$  K. The sample was illuminated by the second harmonic light of the picosecond mode-locked Ti:sapphire laser (pulse width  $\sim 8$  ps, repetition rate 80 MHz). The photons scattered via the RHPS were collected by the lenses into the two optical fibre bundles. The two sets of polarization filters, each of which consists of a quarter-wave plate and a linear polarizer, are put in front of the fibre bundles. The collected photons are filtered by the two independent monochromators and are detected by the two photon-counting PMTs. The correlation of single pulses from the PMTs are recorded by the time-interval analyser with temporal resolution  $\sim 300$  ps. The emission spectrum was observed by the charge-coupled device (CCD) camera. PC (1) and PC (2) are computers. SHG, second-harmonic generation; LBO, lithium borate.

see that  $\hat{\rho}$  has significant amplitudes at  $|HH\rangle\langle HH|$ ,  $|VV\rangle\langle VV|$ ,  $|HH\rangle\langle VV|$  and  $|VV\rangle\langle HH|$ . Because the off-diagonal components,  $|HH\rangle\langle VV|$  and  $|VV\rangle\langle HH|$ , represent the coherence between  $|HH\rangle$  and  $|VV\rangle$  states, the result was that the generated photon pair held the quantum entanglement with good coherence. In fact, assuming that the two-photon state is a maximally entangled state:

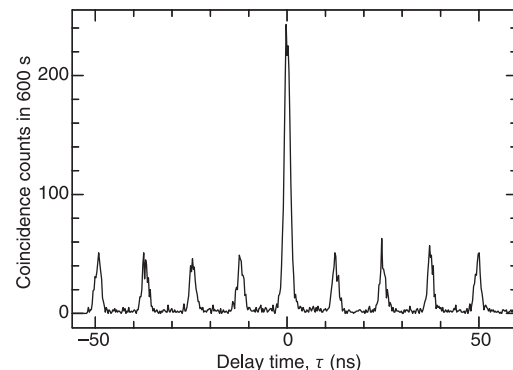
$$|\Psi\rangle = \frac{1}{\sqrt{2}}(|HH\rangle + e^{i\theta}|VV\rangle) \quad (2)$$

we obtained fidelity  $F \equiv \langle \Psi | \hat{\rho} | \Psi \rangle = 0.83$ . This value surpasses the classical limit  $F = 0.5$ , indicating the quantum entanglement of the two-photon polarization state.

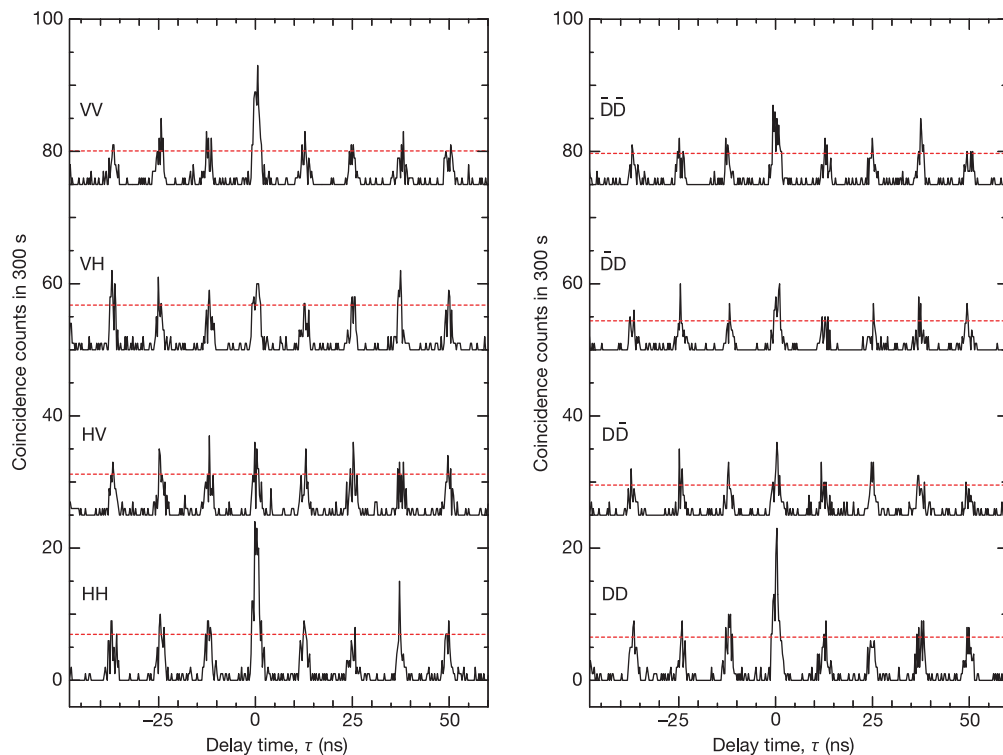
We now consider the degree of entanglement observed in our experiment. Although the observed values of fidelity surpassed the classical limits, they nevertheless were not extremely high. To survey the characteristics of our measured state further we calculated the linear entropy  $S_L$  and the values of tangle  $T$  to evaluate the degrees of disorder and entanglement, respectively<sup>16,17</sup>. The values calculated from the obtained density matrix (Fig. 6) were  $S_L = 0.22$  and  $T = 0.65$ , respectively, indicating that the purity, as well as the degree of entanglement of our state, was somewhat degraded. One possible reason for the partial degradation of entanglement is our experimental geometry; it is known that the light scattered at finite angles from the pump beam is partially polarized<sup>9</sup>, and thus the



**Figure 3** Emission spectrum of the CuCl crystal at 4 K. The photon energy of the pump light (3.1861 eV) is indicated by the downward arrow. The peaks indicated by LEP and HEP are the lights emitted via the RHPS. The peaks  $M_L$  and  $M_T$ , which are not used in our experiment, are resonant biexciton luminescence leaving longitudinal and transverse excitons, respectively.

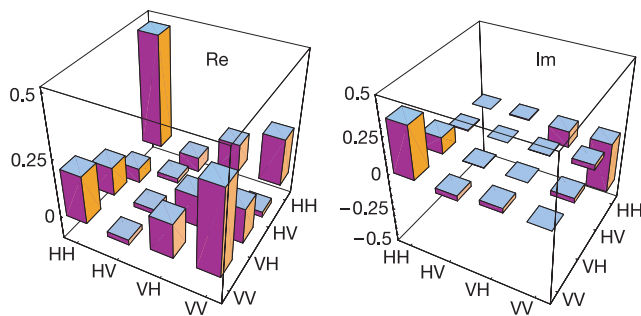


**Figure 4** Photon correlation histogram between the photons emitted via the RHPS without polarization analysers. The central peak at  $\tau = 0$  is a result of the emission of photon pairs from a biexciton. The side peaks originate from uncorrelated pairs of photons emitted from independent biexcitons that different pump pulses generate.



**Figure 5** Photon correlation histogram for linear polarization combinations  $HH$ ,  $HV$ ,  $VH$ ,  $VV$ ,  $DD$ ,  $D\bar{D}$ ,  $\bar{D}D$ , and  $\bar{D}\bar{D}$ , are shown where the first and second letters represent the polarizations of the two photons involved.  $H$ ,  $V$ ,  $D$ , and  $\bar{D}$  refer to horizontal, vertical,  $+45^\circ$  and  $-45^\circ$  polarizations, respectively. The red dotted lines indicate the mean value of the

peaks at  $\tau \neq 0$ , that is, the amount of uncorrelated components originating from two independent biexcitons. Significant enhancement of the peaks at  $\tau = 0$  was observed only in  $HH$ ,  $VV$ ,  $DD$  and  $\bar{D}\bar{D}$  polarizations.



**Figure 6** Graphical representation of the two-photon polarization density matrix reconstructed from the photon correlation measurements for 19 polarization combinations. Shown are the real (Re) and imaginary (Im) parts of the density matrix  $\hat{\rho}$ .

output state differs from the maximally entangled state. However, the expected degree of entanglement, taking into account the effect of our geometry, is  $T = 0.99$ , which is still higher than the observed value. At present, it is not clear whether the additional degradation of entanglement is intrinsic to the RHPS process or was instead caused by another experimental condition. A study to resolve this question is under way.

Thus, we have succeeded in generating entangled photon pairs through RHPS in a semiconductor crystal. To the best of our knowledge, the wavelength ( $\sim 390$  nm) of our entangled photons is the shortest ever realized experimentally. With the use of such entangled photons with short wavelengths, it will be possible to achieve quantum imaging with resolutions much better than in the current technology, which is constrained by the classical diffraction

limit<sup>3,18–20</sup>. We also envision using these entangled pairs to generate entanglement between three or four photons in succeeding parametric down-conversion.

So far, a few attempts to generate entangled photons in semiconductors have been reported. One method is to obtain entangled photons by using cascade photon emission from a biexciton via an exciton in a semiconductor quantum dot<sup>4</sup>. Using this approach, those authors succeeded in observing photon pairs that were correlated classically in time and polarization, but no quantum entanglement has been observed. One of the main reasons for the lack of success is the splitting of exciton states into two polarized states, as induced by the asymmetry of the quantum dot shape<sup>4,5</sup>. In our scheme, however, the sample is a single crystal without such asymmetry. Furthermore, our scheme uses the decay of biexcitons into two exciton-polaritons that are slightly off-resonant from the exciton state. Thus, the two photons are emitted almost concurrently from the biexciton without going through the real exciton as an intermediate state. This is another advantage of our scheme, because it eliminates any decoherence that it might otherwise suffer during the long-lived exciton state. □

Received 17 May; accepted 13 July 2004; doi:10.1038/nature02838.

1. Kwiat, P. G. *et al.* New high-intensity source of polarization-entangled photon pairs. *Phys. Rev. Lett.* **75**, 4337–4341 (1995).
2. Kwiat, P. G. *et al.* Ultrabright source of polarization-entangled photons. *Phys. Rev. A* **60**, R773–R776 (1999).
3. Strekalov, D. V. & Dowling, J. P. Two-photon interferometry for high-resolution imaging. *J. Mod. Opt.* **49**, 519–527 (2002).
4. Santori, C. *et al.* Polarization-correlated photon pairs from a single quantum dot. *Phys. Rev. B* **66**, 045308 (2002).
5. Stace, T. M., Milburn, G. J. & Barnes, C. H. W. Entangled two-photon source using biexciton emission of an asymmetric quantum dot in a cavity. *Phys. Rev. B* **67**, 085317 (2003).
6. Benson, O., Santori, C., Pelton, M. & Yamamoto, Y. Regulated and entangled photons from a single quantum dot. *Phys. Rev. Lett.* **84**, 2513–2516 (1999).
7. Yuan, Z. L. *et al.* Electrically driven single-photon source. *Science* **295**, 102–105 (2002).

8. Ueta, M. *et al.* in *Excitonic Processes in Solids* Ch. 3 (Springer, Berlin, 1986).
9. Itoh, T. & Suzuki, T. Excitonic polariton-polariton resonance scattering via excitonic molecules in CuCl. *J. Phys. Soc. Jpn* **45**, 1939–1948 (1978).
10. Hönerlage, B. *et al.* The dispersion of excitons, polaritons and biexcitons in direct-gap semiconductors. *Phys. Rep.* **124**, 161–253 (1985).
11. Kuwata, M., Mita, T. & Nagasawa, N. Polarization rotation effects associated with the two-photon transition of  $\Gamma_1$  excitonic molecules in CuCl. *Opt. Commun.* **40**, 208–211 (1982).
12. Savasta, S., Martino, G. & Girlanda, R. Entangled photon pairs from the optical decay of biexcitons. *Solid State Commun.* **111**, 495–500 (1999).
13. Shimano, R., Svirko, Y. P., Mysyrowicz, A. & Kuwata-Gonokami, M. Efficient two-photon light amplification by a coherent biexciton wave. *Phys. Rev. Lett.* **89**, 233601 (2002).
14. Saba, M. *et al.* High-temperature ultrafast polariton parametric amplification in semiconductor microcavities. *Nature* **414**, 731–735 (2001).
15. James, D. F. V., Kwiat, P. G., Munro, W. J. & White, A. G. Measurement of qubits. *Phys. Rev. A* **64**, 052312 (2001).
16. White, A. G., James, D. F. V., Munro, W. J. & Kwiat, P. G. Exploring Hilbert space: Accurate characterization of quantum information. *Phys. Rev. A* **65**, 01230 (2001).
17. Munro, W. J., Nemoto, K. & White, A. G. The Bell inequality: a measure of entanglement? *J. Mod. Opt.* **48**, 1239–1246 (2001).
18. Boto, A. N. *et al.* Quantum interferometric optical lithography: exploiting entanglement to beat the diffraction limit. *Phys. Rev. Lett.* **85**, 2733–2736 (2000).
19. Walter, P. *et al.* De Broglie wavelength of a non-local four-photon state. *Nature* **429**, 158–161 (2004).
20. Mitchell, M. W., Lundeen, J. S. & Steinberg, A. M. Super-resolving phase measurements with a multiphoton entangled state. *Nature* **429**, 161–164 (2004).

**Acknowledgements** We thank M. Hasegawa for his help in preparing samples. K.E. is grateful to P. G. Kwiat, M. Kuwata-Gonokami and H. Ishihara for discussions. This work was supported in part by the programme “Strategic Information and Communications R & D Promotion Scheme” of the Ministry of Public Management, Home Affairs, Posts and Telecommunications of Japan.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to K.E. (eda@iec.tohoku.ac.jp).

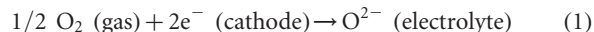
## A high-performance cathode for the next generation of solid-oxide fuel cells

Zongping Shao & Sossina M. Haile

Materials Science, California Institute of Technology, Pasadena, California 91125, USA

Fuel cells directly and efficiently convert chemical energy to electrical energy<sup>1</sup>. Of the various fuel cell types, solid-oxide fuel cells (SOFCs) combine the benefits of environmentally benign power generation with fuel flexibility. However, the necessity for high operating temperatures (800–1,000 °C) has resulted in high costs and materials compatibility challenges<sup>2</sup>. As a consequence, significant effort has been devoted to the development of intermediate-temperature (500–700 °C) SOFCs. A key obstacle to reduced-temperature operation of SOFCs is the poor activity of traditional cathode materials for electrochemical reduction of oxygen in this temperature regime<sup>2</sup>. Here we present Ba<sub>0.5</sub>Sr<sub>0.5</sub>Co<sub>0.8</sub>Fe<sub>0.2</sub>O<sub>3–δ</sub> (BSCF) as a new cathode material for reduced-temperature SOFC operation. BSCF, incorporated into a thin-film doped ceria fuel cell, exhibits high power densities (1,010 mW cm<sup>–2</sup> and 402 mW cm<sup>–2</sup> at 600 °C and 500 °C, respectively) when operated with humidified hydrogen as the fuel and air as the cathode gas. We further demonstrate that BSCF is ideally suited to ‘single-chamber’ fuel-cell operation, where anode and cathode reactions take place within the same physical chamber<sup>3</sup>. The high power output of BSCF cathodes results from the high rate of oxygen diffusion through the material. By enabling operation at reduced temperatures, BSCF cathodes may result in widespread practical implementation of SOFCs.

The primary function of the cathode in a fuel cell based on an oxygen-conducting electrolyte is to facilitate the multistep electrochemical reduction of oxygen<sup>4</sup>:



In a single-chamber fuel cell, the cathode must furthermore be inactive towards oxidation of fuel, and operation at reduced temperatures is essential to minimize undesirable gas-phase reactions<sup>2</sup> (Supplementary Fig. 1). In the present work, we use BSCF as a new cathode material and demonstrate excellent performance in both dual-chamber and single-chamber configurations at temperatures lower than 600 °C. This material, a cubic perovskite (Supplementary Fig. 2) in the BaCoO<sub>3–δ</sub>–SrCoO<sub>3–δ</sub> system, was first developed as a high-temperature (>800 °C) oxygen permeation membrane material<sup>5,6</sup>. Unlike typical cathodes, the A-site cation of BSCF perovskite is an alkaline-earth species rather than a rare-earth element.

We first measured the polarization resistance of BSCF by two-electrode impedance methods using symmetric BSCF|electrolyte|BSCF cells, where the electrolyte was ~1-mm-thick 15 mol% samaria-doped ceria (SDC). As is standard in this field, we use the term area specific resistance (ASR) to describe all resistance terms associated with the electrode, whether they occur at the gas–cathode interface, within the bulk of the cathode, or at the cathode–electrolyte interface. Data were collected under a uniform air atmosphere, both with (three-electrode) and without (two-electrode) a reference electrode; the three-electrode configuration allowed direct measurement of the half-cell voltage with and without an applied direct current. The cathode ASR (Fig. 1a) was determined from raw impedance plots as indicated in Fig. 1b, where the high-frequency offset is due primarily to the electrolyte resistance and the radius of the complete arc to the cathode resistance. The ASR, as determined by both techniques, was remarkably low: 0.055–0.071 Ω cm<sup>2</sup> at 600 °C, and 0.51–0.60 Ω cm<sup>2</sup> at 500 °C. Furthermore, the measured value dropped from ~0.60 Ω cm<sup>2</sup> to ~0.27 Ω cm<sup>2</sup> at 500 °C under an applied current of 0.14 A cm<sup>–2</sup>. Cells fabricated using silver alone as the cathode showed ASR values ~1,000 times larger (Supplementary Fig. 3), demonstrating that silver paste, used to attach silver mesh leads, served only as a current collector. The ASR of the BSCF cathode reported here is substantially lower than that of other single-phase perovskite cathodes measured under similar conditions, and comparable to that of the best composite systems<sup>7–10</sup>.

The performance of the BSCF cathode in a conventional, dual-chamber fuel cell was then investigated, again using SDC as the electrolyte. A thin (20 μm) electrolyte layer was supported on a 700-μm-thick Ni + SDC anode, with a 10–20-μm-thick BSCF cathode layer deposited on the opposing side, after first depositing an additional porous interlayer of SDC (<5 μm in thickness). Air was supplied to the cathode chamber and humidified H<sub>2</sub> (3% H<sub>2</sub>O) to the anode chamber. Peak power densities of ~1,010 mW cm<sup>–2</sup> and 402 mW cm<sup>–2</sup> were obtained at respectively 600 °C and 500 °C (Fig. 2a). These values are more than twice those measured in our laboratory for a similar cell but with SSC + SDC as cathode (Supplementary Fig. 4). In addition to the polarization curves, the cell resistances under open-circuit conditions were measured at various temperatures by impedance spectroscopy. The electrode polarization resistance (the sum of anode and cathode ASRs) is only about 0.021 Ω cm<sup>2</sup> at 600 °C, and 0.135 Ω cm<sup>2</sup> at 500 °C (Fig. 2), amounting to just 14% and 26% of the resistance of electrolyte at these respective temperatures (Supplementary Fig. 5). It is noteworthy that composite SDC + BSCF cathodes, although still very active for oxygen electroreduction, yielded lower power densities than simple BSCF cathodes.

The trilayer fuel cell was further operated in a single-chamber configuration with a propane + O<sub>2</sub> + He mixture in a 4:9:36 volumetric ratio as the feed gas, and a total flow rate of

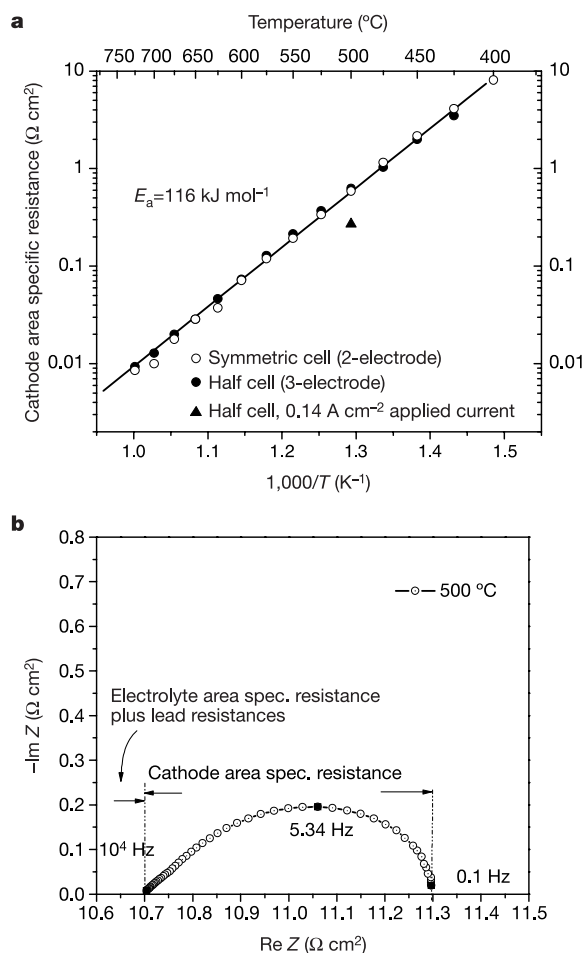


490 ml min<sup>-1</sup>. A peak power density of ~391 mW cm<sup>-2</sup> was observed at a furnace set temperature of 575 °C, with a current density at short circuit of ~1.9 A cm<sup>-2</sup>. At 525 °C the respective values were ~358 mW cm<sup>-2</sup> and ~1.7 A cm<sup>-2</sup>. Upon modifying the BSCF cathode to incorporate 30 wt% SDC, significant improvements in power density were observed in single-chamber mode over simple BSCF. At a furnace set temperature of 500 °C, a peak power density of ~440 mW cm<sup>-2</sup> was achieved (Fig. 3). Hibino *et al.*<sup>3</sup> reported a comparably high peak power density of 403 mW cm<sup>-2</sup> at 500 °C for an electrolyte-supported fuel cell using SSC + SDC as the cathode and ethane as the fuel. However, this cathode was incompatible with propane at temperatures higher than 450 °C (ref. 3). It should be noted that, because of the heat release during partial oxidation at that anode, the real temperature of the single-chamber fuel cell is about 150–245 °C higher than the furnace temperature, depending on the operation conditions (Supplementary Fig. 6). This self-heating in SCFC mode explains the higher power densities obtained from single-chamber than from dual-chamber fuel cells at nominally low temperatures (compare Figs 2a and 3).

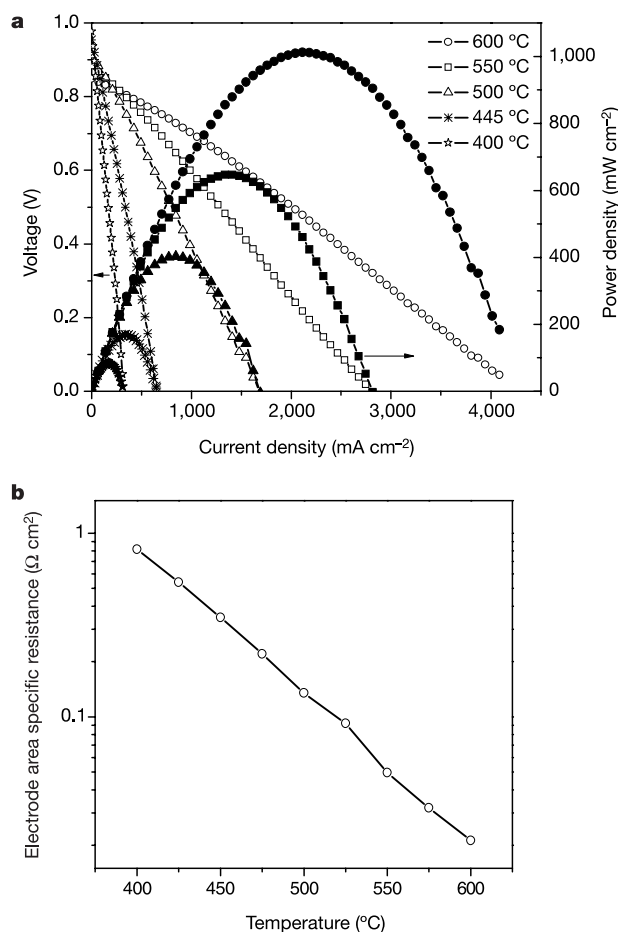
The mechanisms responsible for the excellent performance of

BSCF as a fuel-cell cathode were identified by oxygen permeability measurements (Supplementary Figs 7–9) and extensive impedance spectroscopy studies of symmetric cells (Supplementary Figs 10, 11). The oxygen permeation measurements, combined with thermal gravimetric analysis to determine the oxygen vacancy concentration as a function of oxygen partial pressure, revealed that the oxygen vacancy diffusion rate is  $7.3 \times 10^{-5}$  cm<sup>2</sup> s<sup>-1</sup> at 775 °C and  $1.3 \times 10^{-4}$  cm<sup>2</sup> s<sup>-1</sup> at 900 °C (temperatures at which such measurements could be accurately performed), from 2 to 200 times greater than that of comparable cathode perovskites<sup>11,12</sup>. In addition, the activation energy for oxygen diffusion in BSCF was found to be less than half that for oxygen surface exchange,  $46 \pm 2$  kJ mol<sup>-1</sup> versus  $113 \pm 11$  kJ mol<sup>-1</sup>, suggesting that oxygen surface exchange is the rate-limiting step at low temperatures and that the exceptionally high oxygen diffusivity through BSCF gives it its overall high rate of oxygen electro-oxidation. The oxygen ion conductivity is, in fact, higher than that of SDC. It is for this reason that introduction of a small amount of the electrolyte material decreases cathode performance in dual-chamber configuration, rather than increasing it, as is otherwise observed (for example, in composite LSM + YSZ cathodes<sup>13</sup>).

Impedance spectroscopy of the symmetric cells strongly supported the conclusions of the permeability measurements: specifically, that oxygen diffusion is rapid and surface exchange kinetics are



**Figure 1** The area specific resistance (ASR) of the cathode material BSCF under air, measured both with (three-electrode, half cell) and without (two-electrode, symmetric cell) a reference electrode. **a**, ASR as a function of temperature, and **b**, a typical impedance spectrum, as obtained from the two-electrode, symmetric cell at 500 °C (raw impedance data have been multiplied by the fuel-cell active area, 0.71 cm<sup>2</sup>, so that the cathode ASR corresponds directly to the diameter of the arc associated with the cathode response). The electrolyte, 15 mol% samaria-doped ceria (SDC), is about 1 mm thick, and each electrode about 20 μm thick. The activation energy for the overall cathode resistance is 116 kJ mol<sup>-1</sup>, comparable to that determined for the surface exchange process, indicating that surface exchange at the cathode–air interface is the rate-limiting step.



**Figure 2** Performance obtained from a Ba<sub>0.5</sub>Sr<sub>0.5</sub>Co<sub>0.8</sub>Fe<sub>0.2</sub>O<sub>3-δ</sub> (~20 μm)||Sm<sub>0.15</sub>Ce<sub>0.85</sub>O<sub>2-δ</sub> (~20 μm)||Ni + Sm<sub>0.15</sub>Ce<sub>0.85</sub>O<sub>2-δ</sub> (~700 μm) fuel cell. Air was supplied to the cathode (400 ml min<sup>-1</sup> at STP) and 3% H<sub>2</sub>O-humidified H<sub>2</sub> was supplied to the anode (80 ml min<sup>-1</sup> at STP). **a**, Cell voltage and power density as functions of current density, and **b**, ASRs of the electrodes (sum of anode and cathode contributions) measured under open circuit conditions. The peak power density reaches 1 W cm<sup>-2</sup> at 600 °C.

rate-limiting. In particular, (1) good linearity of the cathode ASR versus reciprocal temperature was observed over the temperature range investigated, 400–725 °C, and the derived activation energy,  $\sim 116 \text{ kJ mol}^{-1}$ , was almost identical to that determined for the oxygen surface exchange step ( $113 \pm 11 \text{ kJ mol}^{-1}$ ). (2) At low temperatures, the cathode ASR was sensitive to the presence of  $\text{CO}_2$  and  $\text{H}_2\text{O}$  in the atmosphere, gases which could only affect surface and not bulk properties. Carbon dioxide increased the interfacial resistance whereas steam decreased it (Supplementary Fig. 10). (3) An increase in the cathode thickness decreased the ASR (without changing the activation energy) (Supplementary Fig. 11), presumably as a result of increasing the area over which surface exchange could occur. (4) The possibility that interfacial charge transfer could be the rate-limiting step is eliminated by the fact that no arc that could be associated with this step appeared in the impedance data.

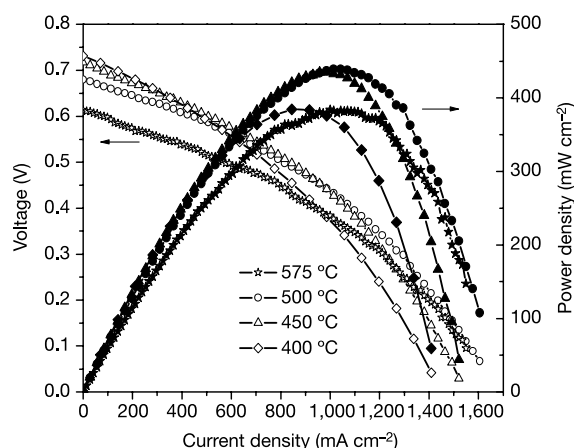
The observation that the apparent cathode ASR drops under applied current (without detectable changes in structure as determined by *ex situ* X-ray diffraction), Fig. 1a, is additionally consistent with a picture in which the surface exchange step is rate-limiting. Specifically, it is known that oxygen adsorption and desorption rates can differ significantly, with desorption occurring much more slowly than adsorption<sup>14</sup>. In the absence of an applied current there is no net ion flux, and thus the adsorption and desorption processes contribute equally to the measured resistance. Under an applied current, however, net adsorption occurs at the electrode functioning as the cathode, and the measured rate is dominated by this faster process, resulting in the observed reduction in ASR. Such an effect could not occur for an overall process limited by bulk diffusion.

A key characteristic of a useful cathode for single-chamber fuel-cell applications is a low activity towards fuel oxidation under the oxidant + fuel environment. It is notable that, together with its high activity of oxygen electroreduction, BSCF exhibits the quality of low activity towards propane oxidation. Under stoichiometric conditions ( $\text{O}_2:\text{C}_3\text{H}_8 = 5:1$  with 95 vol.% helium) and at 500 °C, the propane conversion rates over BSCF, LSCF and SSC are 5.3%, 35.5% and 16.1%, respectively (Fig. 4). The high activity of SSC

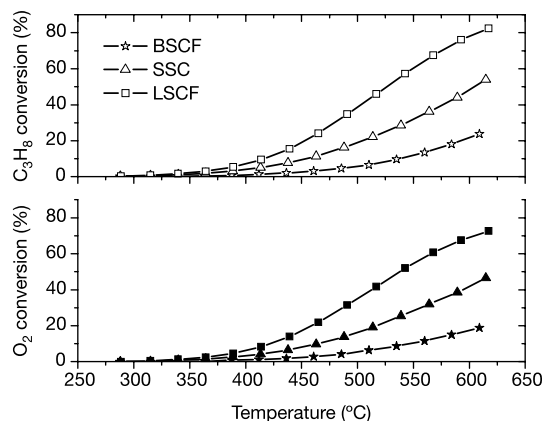
towards propane oxidation is probably responsible for the relatively poor performance of this cathode<sup>3</sup>. Incorporation of electrolyte material into the BSCF cathode under the single-chamber fuel-cell configuration improves performance (unlike the dual-chamber results), and we speculate that the electrolyte is beneficial for limiting the detrimental influence of *in situ* generated  $\text{CO}_2$  on the oxygen surface exchange kinetics of BSCF.

For practical applications, in addition to high power density, good fuel-cell stability is essential. Degradation can occur either by reduction/phase decomposition under low oxygen partial pressures or by phase segregation under an oxygen partial pressure gradient. Whereas phase segregation under a steep oxygen concentration gradient has been reported for BSCF<sup>5,15</sup>, such conditions are unlikely to exist in the fuel-cell cathode because of rapid oxygen diffusion through BSCF. Furthermore, we found that reduction of BSCF occurs only at oxygen partial pressures of  $< 10^{-6}$ – $10^{-8}$  atm. These partial pressures are equivalent to cathode polarization drops of 0.20–0.23 V at 500–600 °C for a cell exposed to ambient air at the cathode. At lower temperatures, polarization losses increase, as does the tendency towards phase segregation, and thus stability at low temperatures provides a stringent measure of long-term viability. A preliminary examination of stability was carried out using a dual-chamber fuel cell with a BSCF-based cathode, operated at 390 °C under an air/3% humidified  $\text{H}_2$  gradient. The cell showed essentially stable performance within the test period of 120 h (Supplementary Fig. 12). Under the fuel-rich, single-chamber conditions employed above, the thermodynamic oxygen partial pressure is as low as  $10^{-28}$  atm at 500 °C, conditions that could well induce the reduction of BSCF. Because such reduction was not observed in our experiments, we conclude that the poor activity of BSCF for propane oxidation results in local oxygen partial pressures in the vicinity of the cathode which are high enough to prevent reduction of cobalt.

The present study demonstrates that very high power densities can be achieved at low temperatures in both dual-chamber and single-chamber fuel cells using  $\text{Ba}_{0.5}\text{Sr}_{0.5}\text{Co}_{0.8}\text{Fe}_{0.2}\text{O}_{3-\delta}$  as a cathode. A significant fraction of the overall cell resistance even at 500 °C ( $\sim 74\%$ ) in dual-chamber configuration is attributable to the 20- $\mu\text{m}$ -thick electrolyte. The SDC used here has a measured conductivity of  $\sim 4.7 \times 10^{-3} \text{ S cm}^{-1}$  at 500 °C (Supplementary Fig. 5), which is lower than the best reported value for doped ceria,  $\sim 9.5 \times 10^{-3} \text{ S cm}^{-1}$  (ref. 16). Thus, a decrease in electrolyte resist-



**Figure 3** Cell voltage and power density as functions of current density obtained in single-chamber mode from a  $\text{Ba}_{0.5}\text{Sr}_{0.5}\text{Co}_{0.8}\text{Fe}_{0.2}\text{O}_{3-\delta} + \text{Sm}_{0.15}\text{Ce}_{0.85}\text{O}_{2-\delta}$  ( $\sim 20 \mu\text{m}$ )  $|\text{Sm}_{0.15}\text{Ce}_{0.85}\text{O}_{2-\delta}$  ( $\sim 20 \mu\text{m}$ )  $|\text{Ni} + \text{Sm}_{0.15}\text{Ce}_{0.85}\text{O}_{2-\delta}$  ( $\sim 700 \mu\text{m}$ ) fuel cell. The BSCF:SDC ratio in the cathode was 70:30 by weight. A mixture of propane, oxygen and helium, flowing (at STP) at  $40 \text{ ml min}^{-1}$ ,  $90 \text{ ml min}^{-1}$  and  $360 \text{ ml min}^{-1}$ , respectively, was used as the feed gas. The temperature reported is that of the furnace; local heating due to partial oxidation at the anode can result in considerably higher temperatures ( $\Delta T \approx 188$ – $240$  °C, Supplementary Fig. 6). The peak power density is relatively insensitive to furnace temperature, broadly peaking at 500 °C at a value of  $\sim 440 \text{ mW cm}^{-2}$ .



**Figure 4** Comparison of the catalytic activities of BSCF ( $\text{Ba}_{0.5}\text{Sr}_{0.5}\text{Co}_{0.8}\text{Fe}_{0.2}\text{O}_{3-\delta}$ ), SSC ( $\text{Sm}_{0.5}\text{Sr}_{0.5}\text{CoO}_{3-\delta}$ ), and LSCF ( $\text{La}_{0.6}\text{Sr}_{0.4}\text{Co}_{0.2}\text{Fe}_{0.8}\text{O}_{3-\delta}$ ) oxides towards propane oxidation under stoichiometric conditions. Test samples were exposed to a mixture of propane, oxygen and helium flowing (at STP) at  $40 \text{ ml min}^{-1}$ ,  $90 \text{ ml min}^{-1}$  and  $360 \text{ ml min}^{-1}$ , respectively. Powders were calcined for 5 h at appropriate temperatures so as to yield comparable specific surface areas: BSCF at 900 °C ( $0.62 \text{ m}^2 \text{ g}^{-1}$ ), SSC at 1,050 °C ( $0.66 \text{ m}^2 \text{ g}^{-1}$ ) and LSCF at 1,100 °C ( $0.65 \text{ m}^2 \text{ g}^{-1}$ ). Low catalytic activity of BSCF is advantageous for single-chamber fuel-cell operation.

ance by a factor of two could realistically be achieved by optimization of its synthesis and composition, or by reduction of its thickness to  $\sim 10\ \mu\text{m}$  (as has already been achieved in the literature<sup>17</sup>), and can be anticipated to result in a peak power density of  $>600\ \text{mW cm}^{-2}$  at  $500^\circ\text{C}$ . Additional increases in power density may be achieved through precise control of the cathode architecture using advanced preparation methods to maximize the surface area over which the oxygen exchange reaction can occur. These steps would probably also yield benefits for operation under single-chamber fuel-cell mode. In addition, a detailed study of possible degradation due to phase segregation induced by the oxygen flux through the cathode, particularly at high current densities, may be necessary for evaluating the long-term viability of BSCF cathodes. □

## Methods

### Cathode powder preparation

Phase-pure  $\text{Ba}_{0.5}\text{Sr}_{0.5}\text{Co}_{0.8}\text{Fe}_{0.2}\text{O}_{3-\delta}$ ,  $\text{Sm}_{0.5}\text{Sr}_{0.5}\text{CoO}_{3-\delta}$  and  $\text{La}_{0.6}\text{Sr}_{0.4}\text{Co}_{0.2}\text{Fe}_{0.8}\text{O}_{3-\delta}$  powders were synthesized by a sol-gel method in which the appropriate metal nitrates were dissolved in water, and a combination of EDTA and citric acid served as complexing agents. Mild heating induced gelation of the solution, and the resulting gel was held at  $250^\circ\text{C}$  for 12 h to remove organics, and then calcined for 5 h under stagnant air at  $900\text{--}1,050^\circ\text{C}$ .

### Fuel-cell fabrication

Electrolyte-supported symmetric cells for impedance studies were prepared by nitrogen-borne spray deposition of BSCF powder suspended in isopropyl alcohol onto both sides of an SDC electrolyte disk ( $\sim 1\ \text{mm}$ ), followed by calcination at  $1,000^\circ\text{C}$  for 5 h under stagnant air. Silver mesh was attached to the electrode surfaces using silver paste as the current collector. Anode-supported thin-film electrolyte fuel cells (trilayer cells) were fabricated using a dual dry-pressing method. The anode, formed from a 60:40 wt% mixture of NiO and SDC powders was dry-pressed into a pellet, and then an electrolyte powder was distributed on the anode surface and co-pressed with the anode. The resultant bilayer was calcined at  $1,350^\circ\text{C}$  for 5 h in air, and then reduced at  $600^\circ\text{C}$  under flowing hydrogen for 5 h. Either simple BSCF or a 70:30 (by weight) BSCF + SDC powder mixture dispersed in isopropyl alcohol was sprayed on the electrolyte surface. The cell was calcined at  $1,000^\circ\text{C}$  for 5 h under flowing nitrogen. The final fuel cells had the following properties: diameter of  $1.33\ \text{cm}$ , anode thickness of  $0.7\ \text{mm}$ , anode porosity of  $\sim 46\%$ , electrolyte thickness of  $\sim 20\ \mu\text{m}$ , cathode thickness of  $10\text{--}20\ \mu\text{m}$ , and cathode porosity of  $\sim 30\%$ .

### Fuel-cell electrochemical characterization

Both symmetric cells and complete trilayer fuel cells were placed in an in-house constructed measurement station, and polarization curves collected using a Keithley 2420 source meter based on the four-probe configuration. In the case of dual-chamber measurements, the trilayer cell was sealed to a quartz tube using silver paste, the cathode side was open to air, and the anode side was exposed to  $3\% \text{H}_2\text{O}$ -humidified  $\text{H}_2$  at a flow rate of  $80\ \text{ml min}^{-1}$  at STP. For the single-chamber fuel-cell test, the whole cell was placed in a quartz tube reactor with an inner diameter of  $15\ \text{mm}$ . Digital mass flow controllers (Aera FC-D980C) were used to introduce propane, oxygen and helium gases to the reactor at various flow rates.

### Electrochemical impedance test

The ASR of BSCF cathode was measured in the two-electrode symmetric cell configuration under air. Impedance measurements were performed using a Solartron 1260A frequency analyser under open circuit conditions from  $10\ \text{mHz}$  to  $10^5\ \text{Hz}$ . The three-electrode half-cell test was conducted by fixing an additional silver electrode to the middle edge of the SDC as a reference electrode, and then measuring the impedance between the working and reference electrodes measured under both open circuit conditions and current polarization conditions using a combined Solartron 1260A frequency response analyser and a PAR EG&G 273A potentiostat/galvanostat. The impedance of the single cell was also measured under asymmetric atmospheres under open circuit conditions.

### Catalytic activity

The catalytic activity of BSCF and other cathode materials towards propane oxidation was tested in a flow-through quartz reactor with a  $6\ \text{mm}$  inner diameter, containing a fixed bed of  $0.3\ \text{g}$  catalyst mixed with  $1.5\ \text{g}$  of silica granules. A gas mixture of  $\text{He} + \text{C}_3\text{H}_8 + \text{O}_2$  was passed through the reactor, which was heated to various temperatures using an electrical furnace. The effluent gas was analysed using an in-line Varian CP-4900 Micro GC.

Received 25 February; accepted 25 June 2004; doi:10.1038/nature02863.

- Steel, B. C. H. & Heinzel, A. Materials for fuel-cell technologies. *Nature* **414**, 345–352 (2001).
- Brandon, N. P., Skinner, S. & Steele, B. C. H. Recent advances in materials for fuel cells. *Annu. Rev. Mater. Res.* **33**, 183–213 (2003).
- Hibino, T. et al. A low-operating-temperature solid oxide fuel cell in hydrocarbon-air mixtures. *Science* **288**, 2031–2033 (2000).
- Fleig, J. Solid oxide fuel cell cathodes: Polarization mechanisms and modeling of the electrochemical performance. *Annu. Rev. Mater. Res.* **33**, 361–382 (2003).
- Shao, Z. P. et al. Investigation of the permeation behavior and stability of a  $\text{Ba}_{0.5}\text{Sr}_{0.5}\text{Co}_{0.8}\text{Fe}_{0.2}\text{O}_{3-\delta}$  oxygen membrane. *J. Membr. Sci.* **172**, 177–188 (2000).

- Shao, Z. P., Dong, H., Xiong, G. X., Cong, Y. & Yang, W. S. Performance of a mixed-conducting ceramic membrane reactor with high oxygen permeability for methane conversion. *J. Membr. Sci.* **183**, 181–192 (2001).
- Huang, K., Wan, J. & Goodenough, J. B. Increasing power density of LSGM-based solid oxide fuel cells using new anode materials. *J. Electrochem. Soc.* **148**, A788–A794 (2001).
- Ishihara, T., Fukui, S., Nishiguchi, H. & Takita, Y. La-doped  $\text{BaCoO}_3$  as a cathode for intermediate temperature solid oxide fuel cells using a  $\text{LaGaO}_3$  based electrolyte. *J. Electrochem. Soc.* **149**, A823–A828 (2002).
- Ralph, J. M., Rossignol, C. & Kumar, R. Cathode materials for reduced-temperature SOFCs. *J. Electrochem. Soc.* **150**, A1518–A1522 (2003).
- Xia, C. R. & Liu, M. L. Novel cathodes for low-temperature solid oxide fuel cells. *Adv. Mater.* **14**, 521–523 (2002).
- Kim, S., Yang, Y. L., Jacobson, A. J. & Abeles, B. Diffusion and surface exchange coefficients in mixed ionic and electronic conducting oxides from the pressure dependence of oxygen permeation. *Solid State Ionics* **106**, 189–195 (1998).
- Xu, S. J. & Thomson, W. J. Oxygen permeation rates through ion-conducting perovskite membranes. *Chem. Eng. Sci.* **54**, 3839–3850 (1999).
- Tsai, T. & Barnett, S. A. Effect of LSM-YSZ cathode on thin-electrolyte solid oxide fuel cell performance. *Solid State Ionics* **93**, 207–217 (1997).
- Fukunaga, H., Koyama, M., Takahashi, N., Wen, C. & Yamada, K. Reaction model of dense  $\text{Sm}_{0.5}\text{Sr}_{0.5}\text{CoO}_3$  as SOFC cathode. *Solid State Ionics* **132**, 279–285 (2000).
- van Veen, A. C., Rebeilleau, M., Farrusseng, D. & Mirodatos, C. Studies on the performance stability of mixed conducting BSCFO membranes in medium temperature oxygen permeation. *Chem. Commun.*, 32–33 (2003).
- Steele, B. C. H. Appraisal of  $\text{Ce}_{1-\gamma}\text{Gd}_\gamma\text{O}_{2-\gamma/2}$  electrolytes for IT-SOFC operation at  $500^\circ\text{C}$ . *Solid State Ionics* **129**, 95–110 (2000).
- Xia, C. R. & Liu, M. L. A simple and cost-effective approach to fabrication of dense ceramic membranes on porous substrates. *J. Am. Ceram. Soc.* **84**, 1903–1905 (2001).

Supplementary Information accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work was funded by the Defense Advanced Research Projects Agency, Microsystems Technology Office. Additional support was provided by the National Science Foundation through the Caltech Center for the Science and Engineering of Materials. Selected oxygen permeability measurements were carried out in the Laboratory of Reaction Engineering and Energy, Institute of Research on Catalysis, CNRS, France, during the visit of Z.P.S. there, hosted by C. Mirodatos.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to S.M.H. (smhaile@caltech.edu).

## The transition to a sulphidic ocean ~ 1.84 billion years ago

Simon W. Poulton<sup>1</sup>, Philip W. Fralick<sup>2</sup> & Donald E. Canfield<sup>1</sup>

<sup>1</sup>Danish Center for Earth System Science, Institute of Biology, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark

<sup>2</sup>Department of Geology, Lakehead University, Thunder Bay, Ontario P7B 5E1, Canada

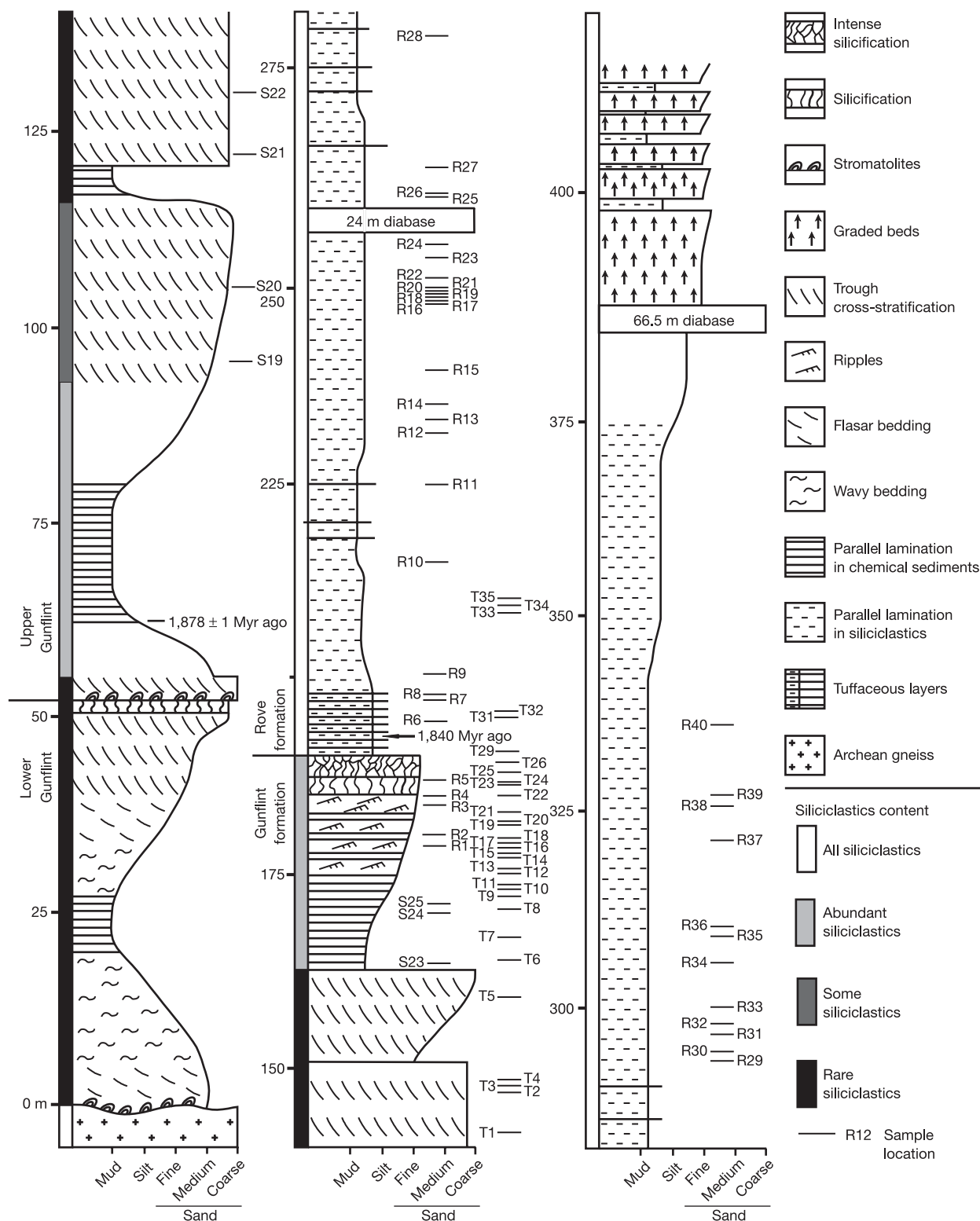
The Proterozoic aeon (2.5 to 0.54 billion years (Gyr) ago) marks the time between the largely anoxic world of the Archean ( $>2.5\ \text{Gyr ago}$ )<sup>1</sup> and the dominantly oxic world of the Phanerozoic ( $<0.54\ \text{Gyr ago}$ ). The course of ocean chemistry through the Proterozoic has traditionally been explained by progressive oxygenation of the deep ocean in response to an increase in atmospheric oxygen around  $2.3\ \text{Gyr ago}$ . This postulated rise in the oxygen content of the ocean is in turn thought to have led to the oxidation of dissolved iron, Fe(II), thus ending the deposition of banded iron formations (BIF) around  $1.8\ \text{Gyr ago}$ <sup>1,2</sup>. An alternative interpretation suggests that the increasing atmospheric oxygen levels enhanced sulphide weathering on land and the flux of sulphate to the oceans. This increased rates of sulphate reduction, resulting in Fe(II) removal in the form of pyrite as the oceans became sulphidic<sup>3</sup>. Here we investigate sediments from the  $\sim 1.8\text{-Gyr-old}$  Animikie group, Canada, which were deposited during the final stages of the main global period of BIF



deposition. This allows us to evaluate the two competing hypotheses for the termination of BIF deposition. We use iron–sulphur–carbon (Fe–S–C) systematics to demonstrate continued ocean anoxia after the final global deposition of BIF and show that a transition to sulphidic bottom waters was ultimately responsible for the termination of BIF deposition. Sulphidic conditions may have persisted until a second major rise in oxygen between 0.8 to

0.58 Gyr ago<sup>4,5</sup>, possibly reducing global rates of primary production and arresting the pace of algal evolution<sup>6</sup>.

The Gunflint and Rove formations of the Animikie group represent a Palaeoproterozoic assemblage of chemical and siliciclastic sedimentary rocks deposited in a shelf setting on the southern margin of Superior province, Ontario, Canada. Depositional models for the Gunflint call for upwelling of anoxic deep ocean



**Figure 1** Stratigraphy of the Gunflint and Rove formations and sample horizons. Samples were collected from unweathered drill core (89-mc-1) and have experienced only very-low-grade metamorphism.

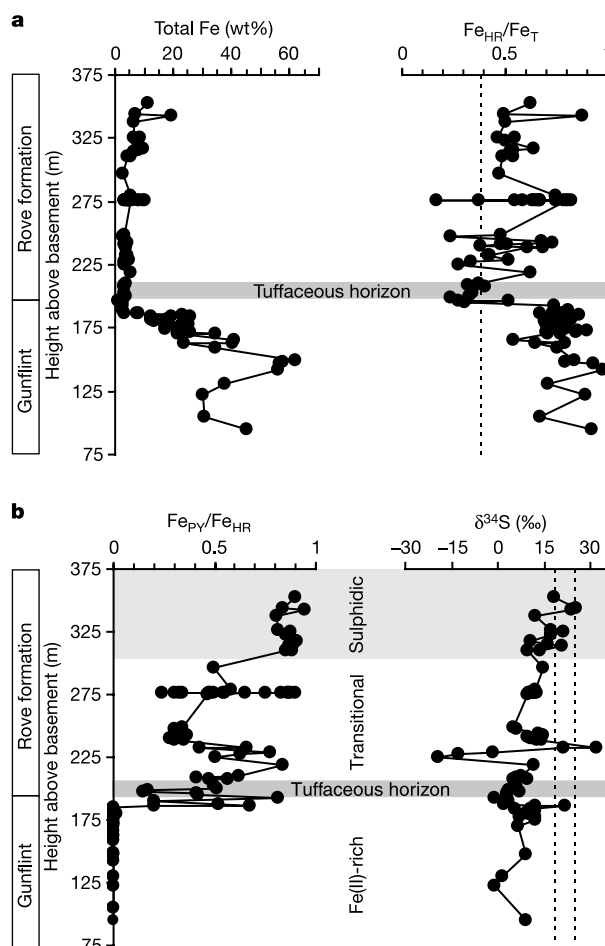
waters containing dissolved Fe(II) onto an oxygenated shelf<sup>7</sup>, with Fe precipitation in outer to mid-shelf areas being partly driven by storm events delivering oxygenated water from the near shore<sup>8</sup>. Rare-earth-element distributions in the Gunflint, combined with Nd isotopic evidence and the limited occurrence of clastics, provide support for a deep-water, probably hydrothermal, source for the Fe<sup>9</sup>. In addition, large Fe fluxes as required to form BIFs of Gunflint magnitude (and similar penecontemporaneous BIFs such as the Frere formation, Western Australia, and the Negaunee iron formation, Michigan, USA) were probably delivered from a major ocean basin<sup>9</sup>. The extensive occurrence of near-shore tidal flat sequences, as found in the Gunflint<sup>10</sup>, also requires unrestricted connection with a large ocean basin. Correlative units to both the Gunflint and Rove formations outcrop over several hundred kilometres to the southwest, representing progressively deeper-water deposits with no evidence of basin closure<sup>8</sup>. Therefore, several lines of evidence point to deposition of the Gunflint and Rove formations in a shelf-slope setting with open connection to the ocean.

Single zircon geochronology on a series of reworked tuffaceous units give an age of  $1,878 \pm 1$  million years (Myr) for the Gunflint<sup>11</sup> (Fig. 1). The Rove formation conformably overlies the Gunflint, but zircon geochronology on tuffaceous material in the basal Rove yields an age of 1,840 Myr (ref. 12) (Fig. 1), indicating a probable major hiatus at the contact.

Iron speciation was investigated to evaluate palaeodepositional redox conditions. The 'indicator of anoxicity' parameter<sup>13</sup> relates highly reactive Fe ( $Fe_{HR}$ ) to total Fe ( $Fe_T$ ), and allows anoxic depositional conditions to be distinguished from oxic conditions.  $Fe_{HR}$  consists of pyrite ( $Fe_{py}$ ) and those Fe phases reactive enough to form pyrite either in the water column or during early sediment diagenesis, specifically magnetite ( $Fe_{mag}$ ), ferric oxides ( $Fe_{ox}$ ) and carbonate minerals ( $Fe_{carb}$ ); thus,  $Fe_{HR} = (Fe_{py} + Fe_{mag} + Fe_{ox} + Fe_{carb})$ . Traditionally  $Fe_{HR}$  has included only  $Fe_{py}$  and  $Fe_{ox}$ , but we have extended  $Fe_{HR}$  to include  $Fe_{mag}$  and  $Fe_{carb}$  (ref. 14), as these phases are important constituents of many Palaeoproterozoic sediments.  $Fe_T$  includes Fe associated with silicate minerals, in addition to the highly reactive Fe phases. In contrast to oxic depositional conditions, modern and ancient sediments deposited beneath anoxic bottom waters usually have  $Fe_{HR}/Fe_T$  ratios exceeding 0.38 (refs 15, 16). High  $Fe_{HR}/Fe_T$  ratios arise from the water column formation of  $Fe_{HR}$ , either as pyrite in sulphidic basins, or magnetite, siderite or ferric oxides in Fe(II)-rich basins, where the additional Fe is largely sourced from anoxic, non-sulphidic parts of the basin<sup>17</sup>.

Total Fe contents are generally high throughout the Gunflint formation, with a gradual decrease towards the top of the unit (Fig. 2a) in association with laminated and silicified horizons (Fig. 1). By contrast, the Rove formation has lower  $Fe_T$ , with concentrations more typical of black shales. Despite these variations in  $Fe_T$ ,  $Fe_{HR}/Fe_T$  ratios generally remain high ( $>0.38$ ) throughout the sequence (Fig. 2a), clearly indicating the persistence of ocean anoxia during deposition of the Rove formation. Relatively low  $Fe_{HR}/Fe_T$  ratios are found in association with rapidly deposited tuffaceous layers and with occasional horizons above this unit (Fig. 2a), and these probably arise owing to dilution of the water column  $Fe_{HR}$  signal by rapidly deposited lithogeneous sediment<sup>17</sup>.

A more detailed examination of Fe speciation can distinguish whether ocean bottom waters were Fe(II)-containing or sulphidic. To achieve this distinction we explore the extent to which  $Fe_{HR}$  has been converted to pyrite (Fig. 2b). Throughout much of the Gunflint formation  $Fe_{py}/Fe_{HR}$  ratios are very low, which, when combined with high  $Fe_{HR}/Fe_T$  ratios (Fig. 2a), is consistent with deposition beneath an anoxic, Fe(II)-containing water column. Towards the top of the Rove formation,  $Fe_{py}/Fe_{HR}$  ratios are persistently very high ( $0.87 \pm 0.04$ ), consistent with nearly complete pyritization of  $Fe_{HR}$ , as found in modern euxinic basins<sup>17</sup>. By contrast, the top 5 m of the Gunflint and lower section of the Rove

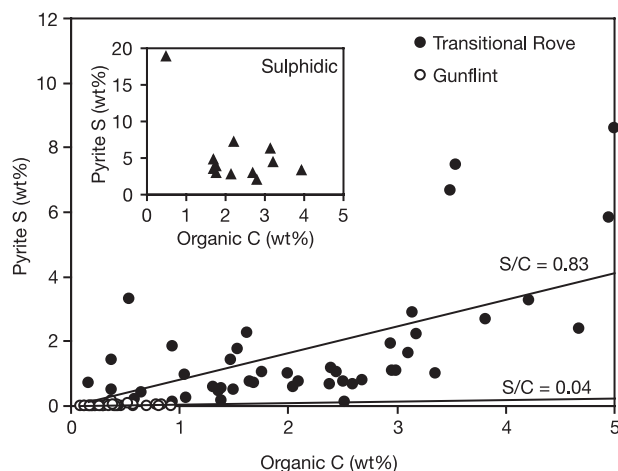


**Figure 2** Iron speciation and S-isotope profiles. **a**, Total Fe and  $Fe_{HR}/Fe_T$ . The dashed line shows an  $Fe_{HR}/Fe_T$  ratio of 0.38; **b**,  $Fe_{py}/Fe_{HR}$  and  $\delta^{34}S$ . Dashed lines represent a probable  $\delta^{34}S$  range of between +18 and +25‰ for late Palaeoproterozoic seawater sulphate<sup>25</sup>. (Full analytical data are available as Supplementary Information).

formation are characterized by intermediate and more variable  $Fe_{py}/Fe_{HR}$  ratios. These could indicate either fluctuating, occasionally sulphidic, water column conditions or variable rates of sediment sulphate reduction.

The Fe speciation data provide compelling evidence for a transition from an anoxic, Fe(II)-containing water column to a sulphidic water column, via a transitional zone corresponding to variable water column or sediment conditions. We additionally note that this trend is not a general feature observed during transitions from BIF to pyritic shale before 1.8 Gyr ago (see Supplementary Information). We must, however, address the possibility that the pyrite enrichment in the upper sediments of the Rove formation resulted from intense sediment sulphate reduction, where  $Fe_{HR}$  was removed from an overlying Fe-containing water column. In this case a diffusional front between sulphide in the sediment and Fe from the water column would be established, probably resulting in the formation of occasional pyrite layers and nodules with changing rates of sediment deposition, for example. The pyrites in the upper Rove, however, are all fine-grained and finely dispersed, arguing against the presence of a diffusional front between Fe and sulphide and supporting the sulphidic water column interpretation.

Relationships between reduced sulphur and organic carbon may also distinguish between deposition in sulphidic, normal oxic marine and sulphate-limited environments<sup>18,19</sup>, although the S/C ratios in Precambrian sediments may not be directly comparable to Phanerozoic counterparts<sup>20,21</sup>. Despite this, the very low S/C ratio of



**Figure 3** S/C ratios for the Gunflint and Rove formations. Inset graph shows data for the uppermost Rove, corresponding to the region for which Fe speciation characteristics imply deposition beneath a sulphidic water column. Four samples directly below the Gunflint/Rove contact with anomalously high pyrite S contents (see Supplementary Information) are not included in the evaluation of the S/C ratio for the Gunflint.

0.04 for the Gunflint (Fig. 3) is entirely consistent with deposition in a low-sulphate environment<sup>20</sup>. In the upper section of the Rove formation, pyrite S is essentially decoupled from organic C and displays a non-zero intercept on the S-axis (Fig. 3). This trend is observed in other middle-Proterozoic sulphidic basins<sup>21</sup> and is attributable to pyrite formation in a sulphidic water column<sup>19</sup>. The lower section of the Rove formation displays a positive correlation between pyrite S and organic C but with considerable scatter ( $r = 0.63$ ), as might be expected in a transitional zone between Fe(II)-rich and sulphidic conditions. Thus, S/C relationships in the Gunflint and Rove formations provide supporting evidence for a transition to euxinic conditions at  $\sim 1.84$  Gyr ago.

The  $\delta^{34}\text{S}$  of pyrite S generally displays a relatively narrow range (Fig. 2b), with occasional horizons in the Gunflint and transitional Rove showing greater variability. In addition, average  $\delta^{34}\text{S}$  values are relatively heavy for each unit (Gunflint,  $8.4 \pm 5.4\text{‰}$ ; transitional Rove,  $8.9 \pm 7.7\text{‰}$ ; sulphidic Rove,  $17.2 \pm 5.1\text{‰}$ ) and show a gradual increase towards the top of the sequence. The narrow range in isotopic composition for the uppermost Rove is consistent with modern and ancient sediments deposited under euxinic conditions<sup>22,23</sup>. The wider range in isotopic fractionations observed in the Gunflint and transitional Rove are similar to those commonly observed during early and late-stage diagenetic pyrite formation<sup>24</sup>. In these horizons, pyrite is mainly associated with mafic pyroclasts or occurs disseminated throughout pore-lining cement, and occasionally pyrite replaces siderite grains<sup>20</sup>. Such morphologies indicate pyrite formation at, or just below, the sediment–water interface, or later-stage diagenetic pyritization in the case of siderite replacement<sup>20</sup>. A thin zone of coarse-grained layered pyrite aggregates is also found towards the top of the transitional Rove. Such layering suggests non-steady-state conditions, such as periodic release of sulphide from sediments into an Fe-rich water column, as discussed above, or possibly the massive precipitation of water column Fe in a transition to sulphidic conditions. The low range in S isotopic fractionations (Fig. 2b) suggests an isotopically homogeneous sulphide source and would tend to support a water column origin for the sulphide.

Estimates for the isotopic composition of seawater sulphate during the late Palaeoproterozoic range from 18‰ to 25‰ (ref. 25). The  $\delta^{34}\text{S}$  of pyrite in the uppermost Rove approaches these values (Fig. 2b), indicating almost complete reduction of water

column sulphate to sulphide. Below this horizon, maximum fractionations relative to seawater sulphate of  $\sim 40\text{‰}$  occur, as observed elsewhere during this time period<sup>6,26</sup>. However, throughout much of the Gunflint and transitional Rove relatively small fractionations from seawater sulphate are observed (Fig. 2b), which, together with low S/C ratios, indicate relatively low sulphate concentrations. Under such conditions, pore-water sulphate is rapidly consumed and sulphide forms with an isotopic composition similar to seawater sulphate<sup>20</sup>. Large fractionations are sometimes preserved, so seawater sulphate concentrations were probably above 200  $\mu\text{M}$ , the minimum value required to induce large fractionations<sup>27</sup>, but less than a maximum value of 2.4 mM implied for the middle Proterozoic<sup>21</sup>.

Taken together, our results suggest continued ocean anoxia with Fe(II)-containing bottom waters after deposition of the Gunflint. This gave way to persistent sulphidic conditions by the upper portion of the Rove formation. Because this sediment package was probably deposited with unrestricted access to the open ocean, widespread sulphidic conditions are indicated. Numerous lines of evidence denote persistent sulphidic conditions within 1.73–1.49-Gyr-old sediments from the McArthur basin, northern Australia<sup>21,28–30</sup>. Although the McArthur seaway was probably an intracratonic basin, Mo isotope constraints indicate widespread ocean anoxia at this time<sup>29</sup>. Further studies are required to determine the precise temporal and spatial extent of such conditions, and their ultimate influence on biological, atmospheric and climatic evolution. □

Received 27 April; accepted 29 July 2004; doi:10.1038/nature02912.

- Holland, H. D. *The Chemical Evolution of the Atmosphere and Oceans* (Princeton Univ. Press, Princeton, 1984).
- Cloud, P. E. A working model of the primitive Earth. *Am. J. Sci.* **272**, 537–548 (1972).
- Canfield, D. E. A new model for Proterozoic ocean chemistry. *Nature* **396**, 450–453 (1998).
- Derry, L. A., Kaufmann, A. J. & Jacobsen, S. B. Sedimentary cycling and environmental change in the Late Proterozoic: Evidence from stable and radiogenic isotopes. *Geochim. Cosmochim. Acta* **56**, 1317–1329 (1992).
- Canfield, D. E. & Teske, A. Late Proterozoic rise in atmospheric oxygen concentrations inferred from phylogenetic and stable isotope studies. *Nature* **382**, 127–132 (1996).
- Anbar, A. D. & Knoll, A. H. Proterozoic ocean chemistry and evolution: A bioinorganic bridge? *Science* **297**, 1137–1142 (2002).
- Drever, J. I. Geochemical model for the origins of Precambrian banded iron formations. *Geol. Soc. Am. Bull.* **85**, 1099–1106 (1974).
- Pfahls, P. K. & Fralick, P. W. Depositional controls on Paleoproterozoic shallow-water iron formation accumulation, Gogebic Range, Wisconsin, U.S.A. *Sedimentology* **54**, 791–808 (2004).
- Derry, L. A. & Jacobsen, S. B. The chemical evolution of Precambrian seawater: Evidence from REEs in banded iron formations. *Geochim. Cosmochim. Acta* **54**, 2965–2977 (1990).
- Ojakangas, R. W. in *Early Proterozoic Geology of the Great Lakes Region* (ed. Medaris, L. G.) *Geol. Soc. Am. Mem.* **160**, 49–66 (1983).
- Fralick, P. W., Davis, D. W. & Kissin, S. A. The age of the Gunflint Formation, Ontario, Canada: single zircon U–Pb age determinations from reworked volcanic ash. *Can. J. Earth Sci.* **39**, 1085–1091 (2002).
- Kissin, S. A., Vallina, D. A., Addison, W. D. & Brumpton, G. R. New zircon ages from the Gunflint and Rove Formations, northwestern Ontario. *Lake Superior Institute Proc.* **49**, 43–44 (2003).
- Raiswell, R., Newton, R. & Wignall, P. B. An indicator of water-column anoxia: Resolution of biofacies variations in the Kimmeridge Clay (Upper Jurassic, U.K.). *J. Sedim. Res.* **71**, 286–294 (2001).
- Poulton, S. W. & Canfield, D. E. Development of a sequential extraction procedure for iron: Implications for iron partitioning in continentally-derived particulates. *Chem. Geol.* (submitted).
- Raiswell, R. & Canfield, D. E. Sources of iron for pyrite formation in marine sediments. *Am. J. Sci.* **298**, 219–245 (1998).
- Poulton, S. W. & Raiswell, R. The low-temperature geochemical cycle of iron: From continental fluxes to marine sediment deposition. *Am. J. Sci.* **302**, 774–805 (2002).
- Anderson, T. F. & Raiswell, R. Sources and mechanisms for the enrichment of highly reactive iron in euxinic Black Sea sediments. *Am. J. Sci.* **304**, 203–233 (2004).
- Berner, R. A. & Raiswell, R. C/S method for distinguishing freshwater from marine sedimentary rocks. *Geology* **12**, 365–368 (1984).
- Raiswell, R. & Berner, R. A. Pyrite formation in euxinic and semi-euxinic sediments. *Am. J. Sci.* **285**, 710–724 (1985).
- Carrigan, W. J. & Cameron, E. M. Petrological and stable isotope studies of carbonate and sulfide minerals from the Gunflint Formation, Ontario: Evidence for the origin of early Proterozoic iron-formation. *Precamb. Res.* **52**, 347–380 (1991).
- Shen, Y., Canfield, D. E. & Knoll, A. H. Middle Proterozoic ocean chemistry: Evidence from the McArthur Basin, Northern Australia. *Am. J. Sci.* **302**, 81–109 (2002).
- Lyons, T. W. Sulfur isotope trends and pathways of iron sulfide formation in upper Holocene sediments of the anoxic Black Sea. *Geochim. Cosmochim. Acta* **61**, 3367–3382 (1997).
- Passier, H. et al. Sulphidic Mediterranean surface waters during Pliocene sapropel formation. *Nature* **397**, 146–149 (1999).
- Schidlowski, M., Hayes, J. M. & Kaplan, I. R. in *Earth's Earliest Biosphere: Its Origin and Evolution* (ed. Schopf, J. W.) 149–186 (Princeton Univ. Press, Princeton, 1983).



25. Strauss, H. The sulfur isotopic record of Precambrian sulfates: new data and a critical evaluation of the existing record. *Precamb. Res.* **63**, 225–246 (1993).
26. Canfield, D. E. & Raiswell, R. The evolution of the sulfur cycle. *Am. J. Sci.* **299**, 697–723 (1999).
27. Habicht, K. H., Gade, M., Thamdrup, B., Berg, P. & Canfield, D. E. Calibration of sulfate levels in the Archean ocean. *Science* **298**, 2372–2374 (2002).
28. Shen, Y., Knoll, A. H. & Walter, M. R. Evidence for low sulphate and anoxia in a mid-Proterozoic marine basin. *Nature* **423**, 632–635 (2003).
29. Arnold, G. L., Anbar, A. D., Barling, J. & Lyons, T. W. Molybdenum isotope evidence for widespread anoxia in mid-Proterozoic oceans. *Science* **304**, 87–90 (2004).
30. Brocks, J. J., Love, G. D., Summons, R. E. & Logan, G. A. Purple sulfur bacteria in an intensely stratified Paleoproterozoic sea. *Geochim. Cosmochim. Acta* **68**( 11S), A796 (2004).

**Supplementary Information** accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank M. Jirsa from the Geological Survey of Minnesota and the staff at the Department of Natural Resources in Ontario, Canada for help in locating and accessing core material. This research was supported by a Marie Curie Individual Fellowship (S.W.P.) and the Danish National Research Foundation (Danish Grundforskningsfond).

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to S.W.P. ([s.poulton@biology.sdu.dk](mailto:s.poulton@biology.sdu.dk)).

## Genetic variation increases during biological invasion by a Cuban lizard

Jason J. Kolbe<sup>1</sup>, Richard E. Glor<sup>1</sup>, Lourdes Rodríguez Schettino<sup>2</sup>, Ada Chamizo Lara<sup>2</sup>, Allan Larson<sup>1</sup> & Jonathan B. Losos<sup>1</sup>

<sup>1</sup>Department of Biology, Campus Box 1137, Washington University, Saint Louis, Missouri 63130-4899, USA

<sup>2</sup>Instituto de Ecología y Sistemática, CITMA, Carretera de Varona km 3.5, Boyeros, La Habana 10800, Apartado Postal 8029, Cuba

A genetic paradox<sup>1,2</sup> exists in invasion biology: how do introduced populations, whose genetic variation has probably been depleted by population bottlenecks, persist and adapt to new conditions? Lessons from conservation genetics show that reduced genetic variation due to genetic drift and founder effects limits the ability of a population to adapt, and small population size increases the risk of extinction<sup>1,3,4</sup>. Nonetheless, many introduced species experiencing these same conditions during initial introductions persist, expand their ranges, evolve rapidly and become invasive. To address this issue, we studied the brown anole, a worldwide invasive lizard. Genetic analyses indicate that at least eight introductions have occurred in Florida from across this lizard's native range, blending genetic variation from different geographic source populations and producing populations that contain substantially more, not less, genetic variation than native populations. Moreover, recently introduced brown anole populations around the world originate from Florida, and some have maintained these elevated levels of genetic variation. Here we show that one key to invasion success may be the occurrence of multiple introductions that transform among-population variation in native ranges to within-population variation in introduced areas. Furthermore, these genetically variable populations may be particularly potent sources for introductions elsewhere. The growing problem of invasive species introductions brings considerable economic and biological costs<sup>5,6</sup>. If these costs are to be mitigated, a greater understanding of the causes, progression and consequences of biological invasions is needed<sup>7</sup>.

Ecological approaches have identified many factors that are important for invasion success and have made quantitative predictions of establishment and spread of invaders<sup>8</sup>. Genetic approaches,

however, have received less attention, even though the loss of genetic variation associated with bottlenecks during introductions may compromise the ability of populations to adapt to new environments<sup>9</sup>, thus limiting their long-term viability. As a result, biological invasions may be evolutionary dead ends<sup>10</sup>. Introduced species are nonetheless pervasive and in some cases evolve rapidly<sup>11</sup>, suggesting an ability of some invasive species to circumvent loss of genetic variation associated with bottlenecks during introductions.

An ideal opportunity to study the genetics of an invading species is provided by the brown anole, *Anolis sagrei*, a small, diurnal lizard introduced worldwide from its native range in the Caribbean. *A. sagrei* is highly invasive; it reaches high population densities, shows exponential range expansion and is competitively superior to and a predator of native lizards<sup>12–14</sup> (Supplementary Fig. S1). Introductions to Jamaica (possibly natural) and southern Florida were first reported in the mid- to late-1800's (ref. 15). The *A. sagrei* invasion of Florida is well documented; it first appeared in the late nineteenth century in the Florida Keys (Supplementary Table S1). For at least half a century its range did not expand appreciably, but beginning in the 1940's, and accelerating in the 1970's, *A. sagrei* spread northward throughout most of Florida (Supplementary Fig. S1). This expansion probably resulted from a combination of northward movement by the early introductions in southern Florida and additional introductions into peninsular Florida<sup>16,17</sup>. More recent introductions have established populations in Hawaii, Grand Cayman, Taiwan, Grenada and other parts of the United States. By examining the geographic distributions of genetic variants in populations today, we infer the history of these introductions and evaluate their consequences for population-genetic variation.

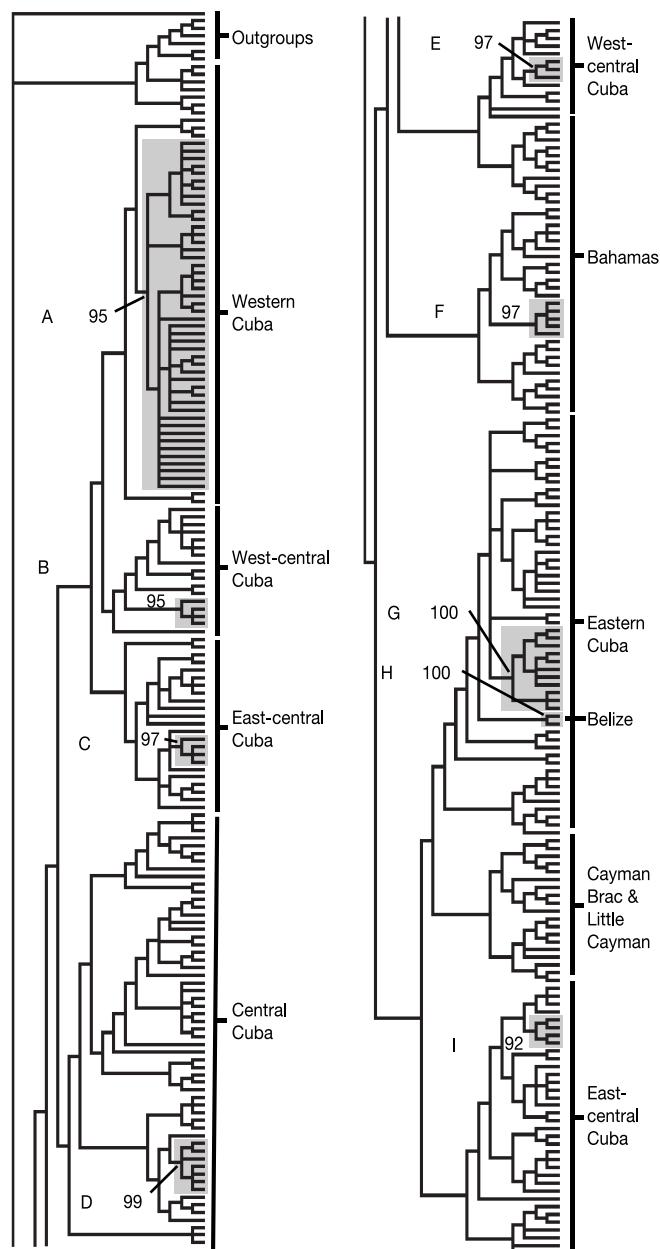
To identify the precise geographic source of colonists, it is essential that the native range be thoroughly sampled. We collected over 600 samples of *A. sagrei* from 71 native and 59 introduced populations to assay levels of genetic variation and to identify the source populations for introductions (Supplementary Table S2). Genetic variation in the native range of *A. sagrei* is highly structured geographically with most genetic variation being among, rather than within, populations (Fig. 1 and Table 1). This genetic structure in the native range allows us to assign haplotypes from introduced populations to geographic localities in the native range with high confidence and to determine the minimum number of introductions. Over 99% of individuals sampled from introduced populations have haplotypes identical or closely related to ones found in Cuba (Fig. 1 and Supplementary Fig. S2). Two kinds of evidence suggest that multiple introductions have occurred into Florida. First, haplotypes from eight different clades are present in Florida, none of which occur together in the native range, indicating eight geographically distinct source populations (Fig. 1). Phylogenetic reconstruction similarly suggests a minimum of eight introductions into Florida. Second, haplotypes whose distributions are distantly separated in Cuba occur together in Florida. Two-thirds of the introduced populations sampled in Florida have haplotypes originating from more than one distinct native population (Fig. 2), but there is no evidence for mixing of haplotypes from different regions within Cuba (Fig. 1 and Table 1). As a result of these multiple introductions from different parts of Cuba, most Florida populations of *A. sagrei* have substantially more genetic variation than native populations. This is evident both in the higher genetic variation within, rather than among, introduced populations (Table 1) as well as their higher mean within-population sequence divergence (Wilcoxon rank sum test:  $z = 3.98$ ;  $P < 0.0001$ ; see Supplementary Fig. S3). Furthermore, more recently established populations have higher genetic variation than older introduced populations owing to the mixture of haplotypes from multiple sources (Figs 2 and 3).

In recent years, *A. sagrei* has been introduced to Hawaii, Grand Cayman, Taiwan and Grenada (Supplementary Table S1), but the

source for these introductions is not documented. As in Florida, these introductions may have occurred one or more times directly from Cuba. Alternatively, owing to the widespread establishment of *A. sagrei* (Supplementary Fig. S1), its high population densities<sup>13</sup> and heightened levels of genetic variation in Florida (Fig. 2), secondary introductions from Florida are possible. Several lines of evidence suggest that Florida populations have been the source for more recent introductions elsewhere. First, recently introduced populations sampled outside Florida contain only haplotypes identical or nearly identical to those sampled in Florida (Supplementary Fig. S2). Second, populations in Hawaii and Taiwan

have combinations of distantly related haplotypes that co-occur in Florida but not in the native range (Fig. 2). Remarkably, some of these secondarily introduced populations have maintained the elevated levels of genetic variation found in Florida populations. These secondary introductions might easily be misinterpreted as resulting from multiple introductions themselves except that the same haplotype combinations exist within Florida populations, and Florida introductions occurred much earlier than those in Hawaii and Taiwan (Fig. 2 and Supplementary Table S1).

By combining genetic variation from multiple source populations, the biological invasion of *A. sagrei* illustrates not only how introduced populations can avoid the reduction in genetic diversity thought to be typical of most introductions, thus solving a paradox in invasive-species genetics<sup>1</sup>, but also how genetic variation within introduced populations can actually increase during an invasion (Fig. 3). Although precedents exist for multiple introductions and increased genetic variation in introduced populations<sup>18,19</sup>, the detailed knowledge of the history of *A. sagrei* introductions and its impact on genetic variability is unparalleled. Populations from initial introductions in the Florida Keys, which today contain haplotypes primarily from western and west-central Cuba, generally have genetic variation similar to or even slightly lower than native Cuban populations (Fig. 2 and Supplementary Fig. S3), probably a result of population bottlenecks during introduction. Conversely, populations from central and northern Florida, which have highly elevated levels of genetic variation, are dominated by haplotypes from central and eastern Cuba, suggesting that additional introductions from Cuba preceded the northward spread of earlier introductions in southern Florida. This pattern is confirmed by the significant geographic genetic structure found in Florida despite high levels of within-population genetic variation (Table 1). Regardless of the chronology or location of each introduction,



**Figure 1** Phylogenetic tree of *A. sagrei*. The tree is from a bayesian analysis of all unique mtDNA haplotypes sampled from the introduced and native ranges of *A. sagrei* (Supplementary Table S2). The tree is split with the top half to the left. Background shading and letters (A–I) indicate nine well-supported clades that are geographically distinct in the native range but contain all haplotypes sampled from introduced populations (see Supplementary Fig. S2). Bayesian posterior probabilities are shown above the branches leading to these nine clades. Geographic regions in the native range are shown to the right.

| Region* (area†)                     | Source of variation‡ | d.f. | % of variation | P-value |
|-------------------------------------|----------------------|------|----------------|---------|
| Native range                        |                      |      |                |         |
| Cuba (100,000 km <sup>2</sup> )     | Among                | 56   | 79.91          | <0.0001 |
|                                     | Within               | 198  | 20.09          |         |
|                                     | Total                | 254  |                |         |
| Bahamas (10,000 km <sup>2</sup> )   | Among                | 8    | 93.22          | <0.0001 |
|                                     | Within               | 34   | 6.78           |         |
|                                     | Total                | 42   |                |         |
| Cayman Brac (35 km <sup>2</sup> )   | Among                | 1    | 57.35          | 0.0059  |
|                                     | Within               | 8    | 42.65          |         |
|                                     | Total                | 9    |                |         |
| Little Cayman (25 km <sup>2</sup> ) | Among                | 1    | 62.57          | 0.0020  |
|                                     | Within               | 9    | 37.43          |         |
|                                     | Total                | 10   |                |         |
| Introduced range                    |                      |      |                |         |
| Florida (70,000 km <sup>2</sup> )   | Among                | 39   | 24.67          | <0.0001 |
|                                     | Within               | 159  | 75.33          |         |
|                                     | Total                | 198  |                |         |
| Jamaica (10,000 km <sup>2</sup> )   | Among                | 7    | 87.67          | <0.0001 |
|                                     | Within               | 27   | 12.33          |         |
|                                     | Total                | 34   |                |         |
| Hawaii (300 km <sup>2</sup> )       | Among                | 3    | 16.98          | 0.1056  |
|                                     | Within               | 9    | 83.02          |         |
|                                     | Total                | 12   |                |         |
| Grand Cayman (200 km <sup>2</sup> ) | Among                | 5    | 17.98          | 0.0430  |
|                                     | Within               | 23   | 82.02          |         |
|                                     | Total                | 28   |                |         |

\* Taiwan and Grenada are not included in these analyses because only one population was sampled in each country.

† The estimated area occupied by *A. sagrei* in each region.

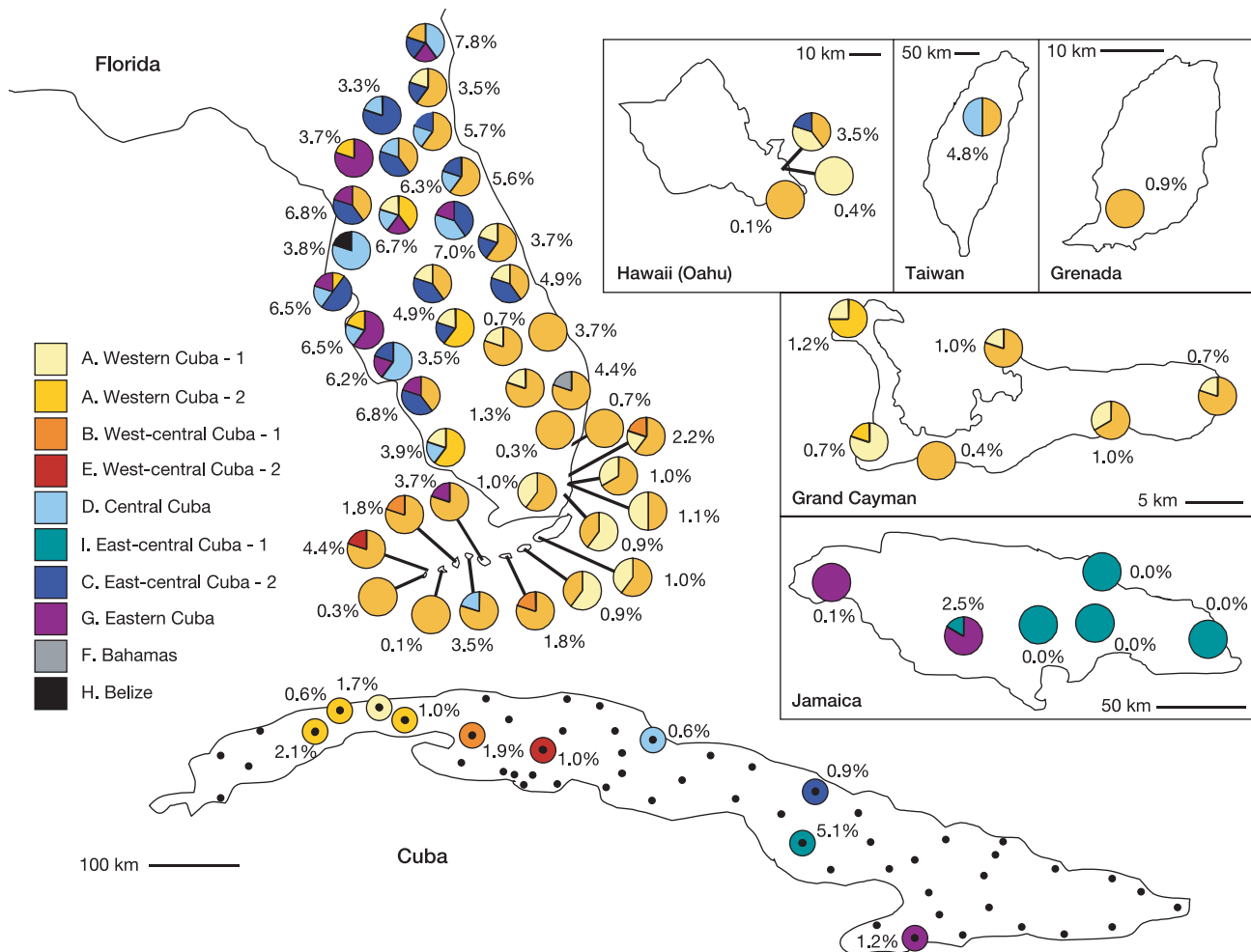
‡ Among populations, within populations or total.

intermingling of genetic variation from distinct introductions produced elevated levels of genetic variation within introduced populations (Figs 2 and 3, and Supplementary Fig. S3).

The lag time between initial Florida introductions in the 1880's and widespread establishment of *A. sagrei* that began in the 1940's is a common feature of biological invasions<sup>2,9,20</sup> (Supplementary Fig. S1). Such delays between the initial establishment of colonists and subsequent expansion are often explained as either an ecological phenomenon, the lag phase in the exponential population growth curve, or an evolutionary phenomenon, the time needed for adaptation to new environments<sup>2</sup>. The genetic data for *A. sagrei* suggest an alternative explanation: Florida populations established during the expansion phase of the invasion contain primarily haplotypes from later introductions, suggesting that separate introductions, not spread of initially introduced populations, explain the range expansion. Unfortunately, the role of multiple introductions in explaining lag times in other biological invasions may be difficult to assess. Detection of multiple introductions may prove difficult except in cases, such as this one, in which geographic sampling of the native range is thorough, genetic variation in native populations is geographically structured and introductions are clearly from multiple areas.

The colonization of *A. sagrei* in Jamaica provides a second example of this phenomenon. As in Florida, populations established by an initial, possibly natural, introduction in the mid-nineteenth century or earlier did not expand rapidly for more than a century<sup>15</sup>, followed by rapid expansion across the island. Genetic data confirm that this rapid expansion resulted from a second introduction from Cuba. Unlike the case in Florida, however, this second introduction came from a single source population; thus, elevated levels of genetic variation occur only in one population in which the two introductions seem to make contact (Fig. 2).

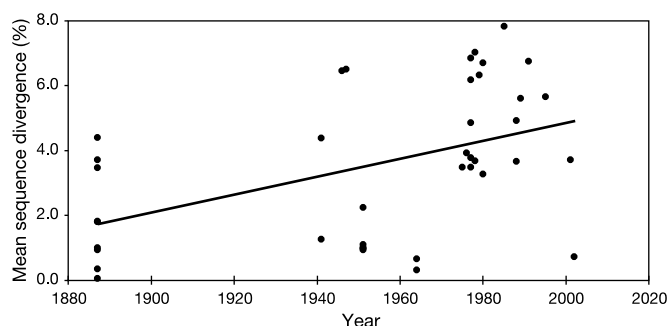
Multiple introductions have clearly enhanced the ability of *A. sagrei* to occupy non-native areas by accelerating its geographic expansion over that possible by simple diffusion alone. Separating the demographic effects of multiple introductions themselves from the effect of increased within-population genetic variation is not easy<sup>2</sup>. Some successful invasions occur despite substantial reductions in genetic variation<sup>21</sup>, so genetic variability is not always necessary, at least in the short term<sup>10</sup>. However, if enhanced genetic variation increases the rate of adaptive evolution, then introduced populations might be expected to show phenotypic divergence from native-range populations. Available data for *A. sagrei* are limited but consistent with this conjecture: Florida populations are more



**Figure 2** Source of genetic variation in introduced *A. sagrei* populations. The map shows the distribution of genetic variation in introduced populations in Florida, Hawaii, Taiwan, Grenada, Grand Cayman and Jamaica, and the source of this genetic variation from native Cuban populations. Black dots indicate native Cuban populations sampled. Colours encircling some black dots denote Cuban populations containing haplotypes to which introduced-population haplotypes are most closely related. Pie charts representing each

introduced population indicate the frequency of haplotypes from different Cuban sources (populations for which only one individual was sampled are not shown). Percentages next to each pie chart give the mean pairwise mtDNA sequence divergence within that population (overall means are 3.5% for Florida and 1.7% for Cuba). Letters (A–I) in the key correspond to clades in Fig. 1 and haplotype networks in Supplementary Fig. S2.





**Figure 3** Genetic variation increases during the biological invasion. The *A. sagrei* invasion of Florida shows a positive relationship between genetic diversity within a population (as measured by mean pairwise percent sequence divergence) and the date that each population was established ( $Y = 0.028X - 50.652$ ;  $P = 0.002$ ;  $R^2 = 0.220$ ). Nearly all introduced Florida populations established after 1975 have an average within-population sequence divergence higher than the mean for Florida populations (overall Florida mean is 3.5%).

variable morphologically than Cuban populations<sup>22</sup>, thus potentially providing variation that could lead to rapid evolution by natural selection. Furthermore, introduced *A. sagrei* in Florida have diverged morphologically from Cuban populations in a variety of characters<sup>22</sup>, including increased body size<sup>23</sup>. Future work is required to examine the underlying genetic basis for the morphological variation observed as well as to investigate its adaptive significance. The role of ecological factors (for example, escape from predators or competitors) also needs to be examined. Studies comparing introduced and native populations as well as introduced populations with varying amounts of genetic variation are needed to disentangle the relative importance and interaction of ecological and genetic factors in determining the spread and adaptive change of this species.

Two disturbing consequences emerge regarding invasive species management and control. First, owing to multiple introductions, invasive populations can become more rather than less genetically variable during an invasion (Figs 2 and 3). Thus, the ability of invasive populations to adapt to new environments will not be constrained for genetic reasons and may be enhanced<sup>24,25</sup>. Second, introduced populations that are widespread and genetically diverse become a source for secondary introductions that maintain elevated genetic variation. Such secondary introductions are likely when the introduced area has higher levels of commercial trade and transport that would aid the spread of a particular species to a greater extent than in the native area; this situation is probably true for Florida versus Cuba. Moreover, if a lack of genetic variation decreases the chances of establishment during an introduction<sup>3,20,26</sup>, both multiple introductions and secondary introductions from highly variable introduced populations may greatly favour the establishment and persistence of new introductions. As a result, invasive-species management, such as biological control, may need to account for genetically diverse biological invasions resulting from multiple introductions<sup>27</sup>. □

## Methods

### Data

We conducted phylogenetic and population-genetic analyses of mitochondrial DNA (mtDNA) sequences from individuals collected in introduced and native populations of *A. sagrei* (Supplementary Table S2). Sampling ranged from one to ten individuals per population with five individuals sampled for most populations. We sequenced an approximately 1,200-base-pair region of mtDNA including the genes encoding ND2, tRNA<sup>Trp</sup> and tRNA<sup>Ala</sup>. We extracted genomic DNA using Viogene extraction kits, amplified gene products using standard PCR methods and purified PCR products using Viogene Gel-M purification kits. Sequencing reactions were run with Big-Dye Terminator Ready-Reaction kits (PerkinElmer) on a Basestation automated sequencer (MJ Research). Sequences were obtained with primers H5730, L4882c and L4437<sup>28</sup>, and aligned manually

using secondary-structural models<sup>28</sup>. All sequences are deposited in GenBank (Supplementary Table S3).

## Phylogenetic and population-genetic analyses

The Bayesian phylogenetic analysis included 313 unique *A. sagrei* sequences and eight outgroup sequences from *Anolis homolechis*, *Anolis bremeri* and *Anolis quadriocellifer*<sup>29</sup>. Likelihood-ratio tests selected the GTR + I +  $\Gamma$  model for the Bayesian analysis<sup>30</sup>, which was run for 1,000,000 generations sampling trees every 9,000 generations<sup>29</sup>. We discarded the trees during the 'burn-in period' (the initial 10% of trees) and made a 50% majority-rule consensus tree from the remaining set of Bayesian trees. We repeated this analysis twice to avoid searching within local optima. This tree was used to identify well-supported clades showing geographic endemism in the native range and containing haplotypes sampled from the introduced range nested within them (Fig. 1). Additionally, we used parsimony to reconstruct the number of separate introductions on each of 102 trees in the Bayesian tree set. Once we had identified clades containing introduced haplotypes, we used statistical parsimony to connect introduced and native-range haplotypes within these clades into haplotype networks using a 95% confidence criterion (Supplementary Fig. S2). The combination of phylogenetic, parsimony reconstruction and population-genetic approaches allowed us to identify populations within the native range from which haplotypes sampled in introduced populations originate.

## Statistical analyses

We tested for significant geographic structure among populations in both the introduced and native ranges using analysis of molecular variance (AMOVA), and for a difference in the average within-population genetic variability between introduced Florida and native Cuban populations using a Wilcoxon rank sum test. Lastly, using approximate dates of introduction or establishment for Florida populations obtained from the literature (see the Florida Fish and Wildlife Conservation Commission's website at <http://wld.fwc.state.fl.us/critters/exotics/SpeciesNumberResults.asp?SPNO=25>), we tested for a relationship between year and genetic diversity within a population (Fig. 3). Also, we tested for an exponential pattern of range expansion within Florida using the same dates of introduction or establishment (Supplementary Fig. S1).

Received 28 April; accepted 30 June 2004; doi:10.1038/nature02807.

1. Allendorf, F. W. & Lundquist, L. L. Introduction: population biology, evolution, and control of invasive species. *Conserv. Biol.* **17**, 24–30 (2003).
2. Sakai, A. K. *et al.* The population biology of invasive species. *Annu. Rev. Ecol. Syst.* **32**, 305–332 (2001).
3. Frankham, R., Ballou, J. D. & Briscoe, D. A. *Introduction to Conservation Genetics* (Cambridge Univ. Press, Cambridge, 2002).
4. Frankham, R. & Ralls, K. Conservation biology: inbreeding leads to extinction. *Nature* **392**, 441–442 (1998).
5. Pimentel, D., Lach, L., Zuniga, R. & Morrison, D. Environmental and economic costs of nonindigenous species in the United States. *Bioscience* **50**, 53–65 (2000).
6. Wilcove, D. S., Rothstein, D., Dubow, J., Phillips, A. & Losos, E. Quantifying threats to imperiled species in the United States. *Bioscience* **48**, 607–615 (1998).
7. Mack, R. N. *et al.* Biotic invasions: causes, epidemiology, global consequences, and control. *Ecol. Appl.* **10**, 689–710 (2000).
8. Kolar, C. S. & Lodge, D. M. Ecological predictions and risk assessment for alien fishes in North America. *Science* **298**, 1233–1235 (2002).
9. Lee, C. E. Evolutionary genetics of invasive species. *Trends Ecol. Evol.* **17**, 386–391 (2002).
10. Queller, D. C. Pax Argentina. *Nature* **405**, 519–520 (2000).
11. Huey, R. B., Gilchrist, G. W., Carlson, M. L., Berrigan, D. B. & Serra, L. Rapid evolution of a geographic cline in size in an introduced population. *Science* **287**, 308–309 (2000).
12. Losos, J. B., Marks, J. C. & Schoener, T. W. Habitat use and ecological interactions of an introduced and a native species of *Anolis* lizard on Grand Cayman, with a review of the outcomes of anole introductions. *Oecologia* **95**, 525–532 (1993).
13. Campbell, T. S. *Analysis of the effects of an exotic lizard (Anolis sagrei) on a native lizard (Anolis carolinensis) in Florida, using islands as experimental units*. PhD thesis, Univ. Tennessee (2000).
14. Gerber, G. P. & Echternacht, A. C. Evidence for asymmetric intraguild predation between native and introduced *Anolis* lizards. *Oecologia* **124**, 599–607 (2000).
15. Williams, E. E. The ecology of colonization as seen in the zoogeography of anole lizards on small islands. *Q. Rev. Biol.* **44**, 345–389 (1969).
16. Campbell, T. S. Northern range expansion of the brown anole (*Anolis sagrei*) in Florida and Georgia. *Herpetol. Rev.* **27**, 155–157 (1996).
17. Lee, J. C. *Anolis sagrei* in Florida: phenetics of a colonizing species I. Meristic characters. *Copeia* **1985**, 182–194 (1985).
18. Novak, S. J. & Mack, R. N. Genetic variation in *Bromus tectorum* (Poaceae): comparison between native and introduced populations. *Heredity* **71**, 167–176 (1993).
19. Facon, B. *et al.* A molecular phylogeography approach to biological invasions of the New World by parthenogenetic Thiarid snails. *Mol. Ecol.* **12**, 3027–3039 (2003).
20. Williamson, M. *Biological Invasions* (Chapman & Hall, London, 1996).
21. Tsutsui, N. D., Suarez, A. V., Holway, D. A. & Case, T. J. Reduced genetic variation and the success of an invasive species. *Proc. Natl Acad. Sci. USA* **97**, 5948–5953 (2000).
22. Lee, J. C. *Anolis sagrei* in Florida: phenetics of a colonizing species III. West Indian and Middle American comparisons. *Copeia* **1992**, 942–954 (1992).
23. Campbell, T. S. & Echternacht, A. C. Introduced species as moving targets: changes in body sizes of introduced lizards following experimental introductions and historical invasions. *Biol. Invasions* **5**, 193–212 (2003).
24. Ellstrand, N. C. & Schierenbeck, K. A. Hybridization as a stimulus for the evolution of invasiveness in plants? *Proc. Natl Acad. Sci. USA* **97**, 7043–7050 (2000).
25. Mooney, H. A. & Cleland, E. E. The evolutionary impact of invasive species. *Proc. Natl Acad. Sci. USA* **98**, 5446–5451 (2001).
26. Ehrlich, P. R. in *Biological Invasions: A Global Perspective* (eds Drake, J. A. *et al.*) 315–328 (Wiley, Chichester, 1989).

27. Roderick, G. K. & Navajas, M. Genes in new environments: genetics and evolution in biological control. *Nature Rev. Genet.* **4**, 889–899 (2003).
28. Macey, J. R., Larson, A., Ananjeva, N. B. & Papenfuss, T. J. Evolutionary shifts in three major structural features of the mitochondrial genome among iguanian lizards. *J. Mol. Evol.* **44**, 660–674 (1997).
29. Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755 (2001).
30. Nylander, J.A.A. MrModeltest v1.0b (Department of Systematic Zoology, Uppsala University, 2002).

Supplementary Information accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank T. Campbell, D. Cardace, P. Colbert, K. de Queiroz, A. Echternacht, J. Gaskin, L. Harmon, J. Knouft, K. Kozak, R. Muller, G. Norval, S. Poe, R. Powell, V. Rivalta Gonzalez, A. Torres Barboza and A. Wright for advice and assistance, and the National Science Foundation and the Environmental Protection Agency Science to Achieve Results (STAR) program for funding.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to J.J.K. (kolbe@biology.wustl.edu).

## Ecosystem stability and compensatory effects in the Inner Mongolia grassland

Yongfei Bai<sup>1</sup>, Xingguo Han<sup>1</sup>, Jianguo Wu<sup>1,2</sup>, Zuozhong Chen<sup>1</sup> & Linghao Li<sup>1</sup>

<sup>1</sup>Laboratory of Quantitative Vegetation Ecology, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China

<sup>2</sup>Faculty of Ecology, Evolution and Environmental Science, School of Life Sciences, Arizona State University, Tempe, Arizona 85287-4501, USA

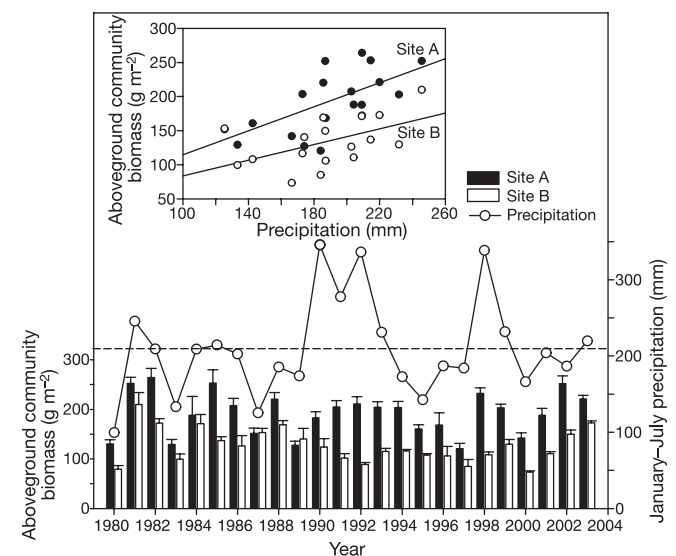
Numerous studies have suggested that biodiversity reduces variability in ecosystem productivity through compensatory effects<sup>1–6</sup>; that is, a species increases in its abundance in response to the reduction of another in a fluctuating environment<sup>1,7</sup>. But this view has been challenged on several grounds<sup>8–10</sup>. Because most studies have been based on artificially constructed grasslands with short duration, long-term studies of natural ecosystems are needed. On the basis of a 24-year study of the Inner Mongolia grassland, here we present three key findings. First, that January–July precipitation is the primary climatic factor causing fluctuations in community biomass production; second, that ecosystem stability (conversely related to variability in community biomass production) increases progressively along the hierarchy of organizational levels (that is, from species to functional group to whole community); and finally, that the community-level stability seems to arise from compensatory interactions among major components at both species and functional group levels. From a hierarchical perspective, our results corroborate some previous findings of compensatory effects<sup>1,4,7,11</sup>. Undisturbed mature steppe ecosystems seem to culminate with high biodiversity, productivity and ecosystem stability concurrently. Because these relationships are correlational, further studies are necessary to verify the causation among these factors. Our study provides new insights for better management and restoration of the rapidly degrading Inner Mongolia grassland.

The role of compensatory interactions between species<sup>6,7</sup> has been a key issue in the debate concerning the diversity–stability relationship of an ecosystem. In particular, because different species respond to environmental fluctuations differently, the reduction in biomass of a certain species is more likely to be recompensed by

the increased biomass of other species in a species-rich rather than species-poor community<sup>1,4,6</sup>. Such compensatory effects have been reported for both plant and animal communities<sup>1,7,11–14</sup>. However, others have argued that plant diversity has no consistent effect, or even a negative effect, on biomass production and ecosystem stability<sup>10,15,16</sup>. Undoubtedly, ecosystem stability depends not only on community composition but also on disturbance, nutrient supply and climatic conditions<sup>3,4,9,13,15,17</sup>, and long-term studies of natural ecosystems are needed for better understanding of compensatory effects and thus the diversity–stability relationship.

Here we present the results of a long-term (1980–2003) study of two natural steppe communities in the Inner Mongolia grassland. The first (site A) is a rhizome-grass-dominated community, and the second (site B) is a bunchgrass-dominated community (see Methods). We classified species into the following five plant functional groups (PFGs) primarily on the basis of life forms: perennial rhizome grass (PR), perennial bunchgrasses (PB), perennial forbs (PF), shrubs and semi-shrubs (SS), and annuals and biennials (AB). PFGs also differ in plant stature, rooting depth, root-to-shoot ratio, water use efficiency, nutrient use efficiency and C:N:P stoichiometry<sup>18–21</sup> (Supplementary Information). Our study addresses the following three questions: first, what are the most important climatic drivers for the aboveground biomass production of steppe communities? Second, how does biomass production respond to precipitation fluctuations at different levels of organization (that is, at the species, plant functional group and community level)? And third, are there detectable compensatory effects reducing the variability in biomass production and thus increasing ecosystem stability?

To address the first question, we used multiple regressions to examine how the aboveground community biomass ( $B_{\text{comm}}$ ) was related to several climatic variables: precipitation (annual, January–July, January–August and May–August); cumulative temperature ( $^{\circ}\text{C}$ ), that is, the accumulated excess when temperature exceeded  $0^{\circ}\text{C}$  (January–July and January–August),  $5^{\circ}\text{C}$  (annual) and  $10^{\circ}\text{C}$



**Figure 1** The relationship between January–July precipitation and total community aboveground biomass ( $B_{\text{comm}}$ ) for the *Leymus chinensis* (site A) and *Stipa grandis* (site B) steppe ecosystems of the Inner Mongolia grassland, using data from 1980 to 2003. Bottom panel:  $B_{\text{comm}}$  was positively correlated to January–July precipitation in site A ( $r^2 = 0.25$ ,  $P = 0.01$ ), but not in site B ( $r^2 = 0.003$ ,  $P = 0.81$ ;  $n = 24$ ). Error bars represent s.e.m., and the horizontal dashed line is the mean January–July precipitation from 1980 to 2003. Top panel: a significant positive correlation was found between  $B_{\text{comm}}$  and January–July precipitation in both sites after removing the four extraordinarily wet years (1990, 1991, 1992 and 1998). For site A (black dots)  $r^2 = 0.49$ ,  $P < 0.001$ ,  $n = 19$ ; for site B (open circles)  $r^2 = 0.35$ ,  $P < 0.01$ ,  $n = 19$ .

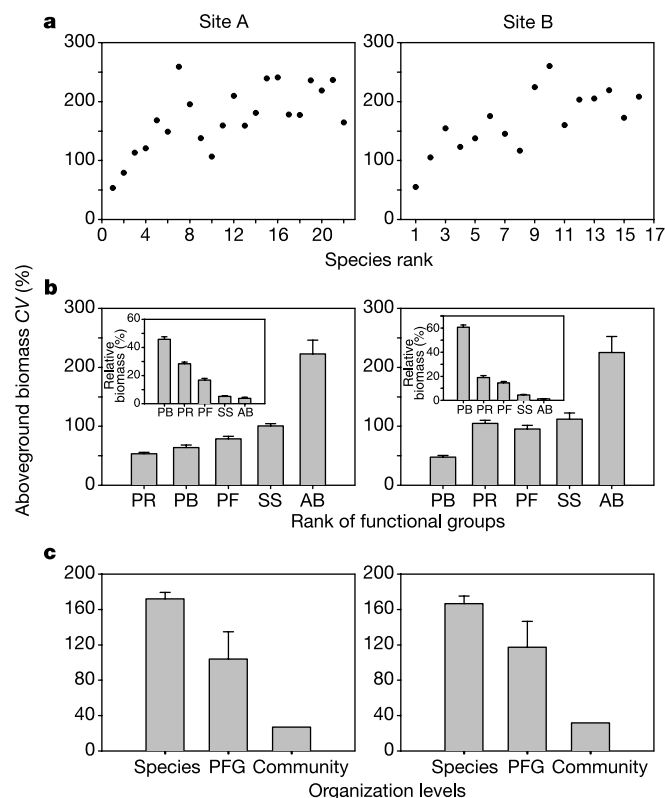
(annual); and the number of days on which temperature-dependent plant growth occurred ( $\geq 5^{\circ}\text{C}$  growing degree days and  $\geq 10^{\circ}\text{C}$  growing degree days). For site A, only January–July precipitation remained as a significant variable in simplified multiple regressions obtained by backward elimination ( $r^2 = 0.25$ ,  $P < 0.01$ ), whereas no significant variable was found for site B. This lack of statistical significance was due mainly to the influence of four extraordinarily wet years (1990, 1991, 1992 and 1998) to which community biomass did not respond proportionately (Fig. 1). When these four years were excluded, January–July precipitation became a highly significant variable for both site A ( $r^2 = 0.49$ ,  $P < 0.001$ ) and site B ( $r^2 = 0.35$ ,  $P < 0.01$ ), and was highly correlated with  $B_{\text{comm}}$  (Fig. 1). January–July precipitation alone explained 35–49% of the variation in biomass for the two sites.

To address the second question, we compared the coefficient of variation in biomass at the levels of individual species ( $CV_{\text{sp.}}$ ), plant functional group ( $CV_{\text{PFG}}$ ) and the whole community ( $CV_{\text{comm}}$ ) (Fig. 2). We calculated  $CV_{\text{sp.}}$  for 22 commonly found species in site A and 16 in site B, and the values ranged from 53.56% to 259.06% for site A and from 55.12% to 260.17% for site B. In both sites, species with higher relative biomass tended to have lower  $CV_{\text{sp.}}$  (Fig. 2a). The coefficient of variation in aboveground biomass varied significantly among certain PFGs in both sites ( $F = 38.51$ ,  $P < 0.0001$  for site A;  $F = 23.28$ ,  $P < 0.0001$  for site B). Specifically, the values of  $CV_{\text{PFG}}$  were significantly different among PR, SS and AB ( $P < 0.01$ ) in site A and among PB, PR and AB in site B (Fig. 2b).  $CV_{\text{PFG}}$  was negatively related to the relative biomass of PFGs ( $B_{\text{rel}}$ ) in both site A

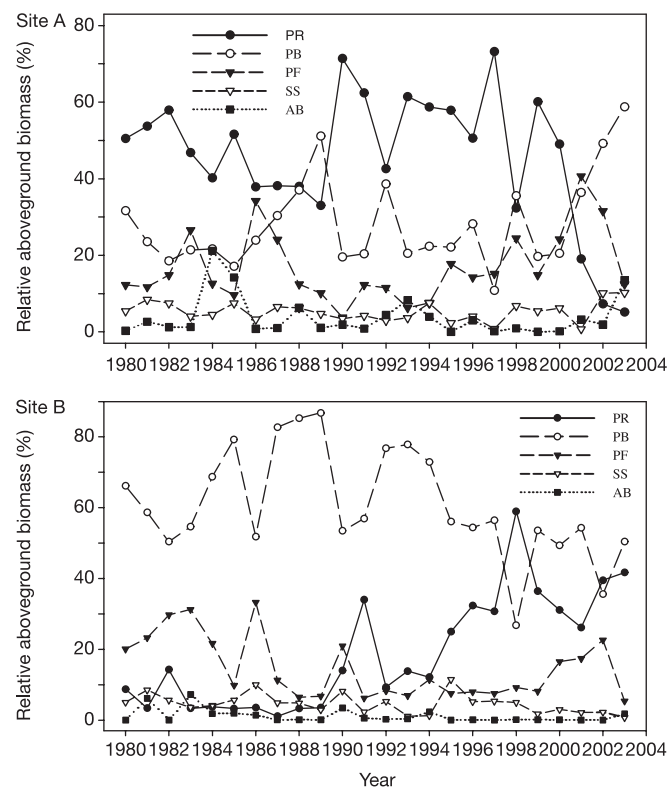
( $CV_{\text{PFG}} = -2.77B_{\text{rel}} + 159.34$ ;  $r^2 = 0.44$ ,  $P < 0.001$ ) and site B ( $CV_{\text{PFG}} = -2.03B_{\text{rel}} + 157.7$ ;  $r^2 = 0.45$ ,  $P < 0.001$ ).

At the community level, the aboveground biomass ( $B_{\text{comm}}$ ) was significantly different between the two sites over the 24-yr period ( $F = 112.99$ ,  $P < 0.0001$ ,  $n = 120$ ). The average  $B_{\text{comm}}$  was  $192.51 \text{ g m}^{-2}$  (s.e.m. = 13.92,  $n = 5$ ) for site A and  $127.04 \text{ g m}^{-2}$  (s.e.m. = 10.14,  $n = 5$ ) for site B. However, the coefficient of variation in  $B_{\text{comm}}$  was not significantly different between the two sites ( $CV_{\text{comm}} = 27.06\%$  for site A and  $31.92\%$  for site B;  $P > 0.05$ ). The values of  $CV_{\text{comm}}$  for these Inner Mongolia steppes are comparable to those for the North America grasslands (25–35%)<sup>22,23</sup>, but much lower than those for semi-arid African grasslands (60–70%)<sup>24</sup>. Ecosystem stability, measured as the coefficient of variation in aboveground biomass<sup>4,10</sup>, increased systematically from the species to the PFG to the community level (that is,  $CV_{\text{sp.}} > CV_{\text{PFG}} > CV_{\text{comm}}$ ) for both site A ( $F = 16.09$ ,  $P < 0.0001$ ) and site B ( $F = 13.71$ ,  $P < 0.001$ ) (Fig. 2c). Such ‘metastability’ has been widely observed in complex hierarchical systems whose apparent stability hinges on the more transient dynamics of their components<sup>25,26</sup>.

Were compensatory effects responsible for the increased stability at the community level? To address this question, we assumed that if an ecosystem in a changing environment had compensatory mechanisms, at least some of its major components (species or PFGs) would show negative correlations in terms of biomass production over time. At the species level, correlation analyses for 231 species pairs in site A and 120 pairs in site B showed that a negative correlation existed for 10 pairs in site A and 15 pairs in site B, whereas a positive correlation was found for 19 pairs in site A and 24 pairs in site B (Supplementary Information). Importantly, the negative correlation existed between dominant and sub-dominant species as well as between dominant and non-dominant species at both sites. In contrast, the positive correlation was found mainly between sub-dominant and non-dominant species or between non-



**Figure 2** Coefficients of variation ( $CV$ s) in aboveground biomass at different organizational levels in the two study sites. **a**, Scatter-plots of species-level  $CV$ s against relative biomass-based species rank in which only common species were included. **b**, Relationship between  $CV$ s of plant functional groups (PFGs) and relative biomass-based ranks of PFGs, with insets showing the average relative biomass values of PFGs in descending order. PR, perennial rhizome grass; PB, perennial bunchgrasses; PF, perennial forbs; SS, shrubs and semi-shrubs; and AB, annuals and biennials. **c**, Comparison of average  $CV$ s at the species, PFG and community levels (means differ significantly at  $P < 0.0001$  in both sites). Error bars in **b** and **c** represent s.e.m.



**Figure 3** Time series of the relative aboveground biomass of plant functional groups from 1980 to 2003. Abbreviations are as in Fig. 2.



Table 1 Correlation coefficients\* between plant functional groups in terms of aboveground biomass

| Plant functional groups† | Site A                      |        | Site B                      |         |
|--------------------------|-----------------------------|--------|-----------------------------|---------|
|                          | Correlation coefficient (r) | P      | Correlation coefficient (r) | P       |
| PR vs PB                 | -0.2673                     | 0.0032 | -0.4843                     | <0.0001 |
| PR vs PF                 | -0.3019                     | 0.0008 | -0.3312                     | 0.0002  |
| PR vs SS                 | 0.1648                      | 0.0721 | -0.2989                     | 0.0010  |
| PR vs AB                 | 0.1616                      | 0.0778 | -0.3189                     | 0.0004  |
| PB vs PF                 | 0.1297                      | 0.1580 | 0.0747                      | 0.4197  |
| PB vs SS                 | 0.2302                      | 0.0114 | 0.2516                      | 0.0058  |
| PB vs AB                 | 0.2668                      | 0.0032 | 0.3001                      | 0.0009  |
| PF vs SS                 | 0.0690                      | 0.4537 | 0.0770                      | 0.4050  |
| PF vs AB                 | -0.2452                     | 0.0069 | 0.2812                      | 0.0020  |
| SS vs AB                 | 0.2233                      | 0.0142 | 0.1587                      | 0.0848  |

\* Spearman's rank correlation coefficients.

† PR, perennial rhizome grass; PB, perennial bunchgrasses; PF, perennial forbs; SS, shrubs and semi-shrubs; and AB, annuals and biennials.

dominant species themselves. A similar trend was found at the PFG level. In site A, a negative correlation existed between PR and PB, between PR and PF, and between PF and AB, whereas a significant positive correlation was found between PB and SS, between PB and AB, and between SS and AB (Table 1). In site B, there was a significant negative correlation between PR and PB, PR and PF, PR and SS, and PR and AB, and a significant positive correlation between PB and SS, PB and AB, and PF and AB (Table 1 and Supplementary Information).

The observation that certain PFGs and species were negatively correlated suggests that compensatory effects may take place at both levels, but primarily among dominant components (that is, species and PFGs with high relative biomass). For example, the perennial rhizome grass, *Leymus chinensis*, alone made up 44.85% of  $B_{\text{comm}}$  in site A, and perennial bunchgrasses made up 61.6% of  $B_{\text{comm}}$  in site B (Fig. 3). Positive correlations at the species and PFG levels, a possible indication of the existence of complementary effects<sup>4,6</sup>, could increase temporal ecosystem variation. Most species pairs (87.4% in site A and 67.5% in site B) did not show any statistically significant correlations in terms of biomass production. This may imply that at the species level a statistical averaging effect might have been operating. But because of the small relative biomass of the species or the PFGs involved (Fig. 3), these effects did not seem significant. The stabilizing effects of the compensatory interactions were also corroborated by the relatively high drought resistance and resilience of these steppe communities. The average value of drought resistance ( $(B_{\text{drought}} - B_{\text{pre-drought}})/B_{\text{pre-drought}}$ ) for five significantly drier years (1980, 1983, 1987, 1995 and 2000 when January–July precipitation was 50–80% of the mean) was -0.27 for site A and -0.22 for site B. These values were higher than those for species-rich ( $\geq 11$  species) experimental grassland plots in North America (about -0.35)<sup>11</sup>. The average values of ecosystem resilience ( $B_{\text{post-drought}}/B_{\text{pre-drought}}$ ) for the five dry years were 1.19 for site A and 1.22 for site B, slightly higher than those for species-rich experimental grassland plots in North America (about 1.1)<sup>11</sup>.

On the basis of a short-term (3 yr) field manipulative drought experiment with artificial grassland plots, it has been suggested that there is a potential trade-off between ecosystem production and stability with increasing species richness<sup>10</sup>. Our results indicate that plant communities in the Inner Mongolia grassland achieve high species richness, productivity and ecosystem stability simultaneously at the late successional stage<sup>18,27,28</sup> (Supplementary Information). These findings shed new light on the effects of precipitation fluctuations on the structure and functioning of the steppe ecosystems, and have important implications for understanding future impacts of climate change, and improving current grassland management practices. For example, as the precipitation regime and dry period pattern are altered, a shift in dominance of species and functional groups may occur. For restoring the vast areas of degraded grasslands in Inner Mongolia<sup>29</sup>, it is important to establish and maintain grassland communities with a high diversity

of dominant species and functional groups, such that compensatory mechanisms can enhance long-term ecosystem productivity and stability in the face of eternal climatic fluctuations (Supplementary Information). □

## Methods

### Study sites and field sampling

The two study sites, each 500 m by 500 m, are the permanent field sites of the Inner Mongolia Grassland Ecosystem Research Station (IMGERS), located in the Xilin River Basin, Inner Mongolia Autonomous Region, China (116°42' E, 43°38' N) and administered by the Institute of Botany, the Chinese Academy of Sciences, Beijing. The growing season in the Inner Mongolia grassland runs from early April to late September for perennial plant species, whereas annual plants usually germinate in early July following the rains. Site A was dominated by a perennial rhizome grass, *Leymus chinensis*, whereas site B was dominated by a perennial bunchgrass, *Stipa grandis*. These two community types represent the most widely distributed grassland communities in the Eurasia steppe region<sup>18,27,28,30</sup>, which is the largest contiguous grassland area in the world. Both sites have been fenced-off since 1979, preventing grazing by large animals. At the time of enclosure, both sites were considered to be in excellent condition, representative of undisturbed, climax steppe communities<sup>28,30</sup>. In 1980, site A and site B had 86 and 61 vascular plant species, respectively, and neither of them changed in species richness over the 24-yr period.

At each site, an east–west transect of 200 × 100 m was established with five equal-sized replicate blocks (40 × 100 m each). Aboveground biomass was sampled during 28–30 August each year by clipping all plants within a 1 × 1 m<sup>2</sup> quadrat that was randomly located within each block. All living vascular plants were sorted into species, dried and weighed. The dry mass of all plant species per quadrat averaged over the five blocks was used to estimate the aboveground community biomass. Because the standing crop of these steppe communities reached the annual peak at the end of August, our estimated community biomass approximated the aboveground net primary productivity of these ecosystems. All meteorological data were obtained from the weather stations of IMGERS.

### Data analysis

Statistical analyses were performed using SAS version 8.0. Analysis of variance (ANOVA) and general linear models were used for analysis of variance. The coefficient of variation was calculated as:  $CV = (\text{standard deviation}/\text{mean}) \times 100$ . Spearman's rank correlation was used to analyse possible correlations among species and PFGs in terms of their aboveground biomass. Multiple regressions and backwards elimination were used to determine the effects of climatic variables on the total community biomass. At the species level, only commonly found species (that is, those present in at least 17 out of the 24 years) were considered. The biomass of the rare species excluded from the analysis was deemed trivial in relation to the total community biomass.

Received 16 May; accepted 13 July 2004; doi:10.1038/nature02850.

1. Naeem, S. & Li, S. Biodiversity enhances ecosystem reliability. *Nature* **390**, 507–509 (1997).
2. Yachi, S. & Loreau, M. Biodiversity and ecosystem productivity in a fluctuating environment: the insurance hypothesis. *Proc. Natl Acad. Sci. USA* **96**, 1463–1468 (1999).
3. Tilman, D. et al. The influence of functional diversity and composition on ecosystem processes. *Science* **277**, 1300–1302 (1997).
4. Tilman, D. The ecological consequences of changes in biodiversity: a search for general principles. *Ecology* **80**, 1455–1474 (1999).
5. Loreau, M. Biodiversity and ecosystem functioning: a mechanistic model. *Proc. Natl Acad. Sci. USA* **95**, 5632–5636 (1998).
6. Loreau, M. Biodiversity and ecosystem functioning: recent theoretical advances. *Oikos* **91**, 3–17 (2000).
7. McNaughton, S. J. Diversity and stability of ecological communities: A comment on the role of empiricism in ecology. *Am. Nat.* **111**, 515–525 (1977).
8. Givnish, T. J. Does diversity beget stability? *Nature* **371**, 113–114 (1994).
9. Huston, M. A. Hidden treatments in ecological experiments: re-evaluating the ecosystem function of biodiversity. *Oecologia* **110**, 449–460 (1997).
10. Pfisterer, A. B. & Schmid, B. Diversity-dependent production can decrease the stability of ecosystem functioning. *Nature* **416**, 84–86 (2002).
11. Tilman, D. & Downing, J. A. Biodiversity and stability in grasslands. *Nature* **367**, 363–365 (1994).

12. Brown, J. H., Whitham, T. G., Ernest, S. K. M. & Gehring, C. A. Complex species interactions and the dynamics of ecological systems: long-term experiments. *Science* **293**, 643–650 (2001).
13. Wardle, D. A., Bonner, K. I. & Barker, G. M. Stability of ecosystem properties in response to aboveground functional group richness and composition. *Oikos* **89**, 11–23 (2000).
14. McNaughton, S. J. Ecology of a grazing ecosystem: The Serengeti. *Ecol. Monogr.* **55**, 259–294 (1985).
15. McCann, K. S. The diversity–stability debate. *Nature* **405**, 228–233 (2000).
16. Huston, M. A. *et al.* No consistent effect of plant diversity on productivity. *Science* **289**, 1255 (2000).
17. Hooper, D. U. & Vitousek, P. M. The effects of plant composition and diversity on ecosystem processes. *Science* **277**, 1302–1305 (1997).
18. Chen, Z. Z. & Wang, S. P. *Typical Steppe Ecosystems of China* (Science, Beijing, 2000).
19. Chen, S. H., Zhang, H., Wang, L. Q., Zhanbula & Zhao, M. L. *Root Systems of Grassland Plants in Northern China* (Jilin Univ. Press, Changchun, 2001).
20. Inner Mongolia Grassland Ecosystem Research Station, Chen, Z. Z., Huang, D. H. & Zhang, H. F. *Research on Grassland Ecosystems* No. 1 112–131 (Science, Beijing, 1985).
21. Bai, Y. F. Influence of seasonal distribution of precipitation on primary productivity of *Stipa krylovii* community. *Acta Phytocol. Sin.* **23**, 155–160 (1999).
22. Knapp, A. K. & Smith, M. D. Variation among biomes in temporal dynamics of aboveground primary production. *Science* **291**, 481–484 (2001).
23. Knapp, A. K. *et al.* Rainfall variability, carbon cycling and plant species diversity in a mesic grassland. *Science* **298**, 2202–2205 (2002).
24. O'Connor, T. G., Haines, L. M. & Snyman, H. A. Influence of precipitation and species composition on phytomass of a semi-arid African grassland. *J. Ecol.* **89**, 850–860 (2001).
25. Wu, J. & Levin, S. A. A spatial patch dynamic modeling approach to pattern and process in an annual grassland. *Ecol. Monogr.* **64**, 447–464 (1994).
26. Wu, J. & Loucks, O. L. From balance of nature to hierarchical patch dynamics: a paradigm shift in ecology. *Q. Rev. Biol.* **70**, 439–466 (1995).
27. Inner Mongolia Grassland Ecosystem Research Station, Li, B., Yong, S. P. & Li, Z. H. *Research on Grassland Ecosystem* No. 3 84–183 (Science, Beijing, 1988).
28. US National Research Council, Wu, J. & Loucks, O. L. *Grasslands and Grassland Sciences in Northern China* 67–84 (National Academy Press, Washington DC, 1992).
29. Tong, C., Wu, J., Yong, S. P., Yang, J. R. & Yong, W. Y. A landscape-scale assessment of steppe degradation in the Xilin River Basin, Inner Mongolia, China. *J. Arid Environ.* **59**, 133–149 (2004).
30. Inner Mongolia Grassland Ecosystem Research Station & Jiang, S. *Research on Grassland Ecosystem* No. 1 1–11 (Science, Beijing, 1985).

**Supplementary Information** accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank D. Huang, Y. Li, Q. Pan, Q. Qi, S. Wang, Yanfen Wang, Yifeng Wang, Z. Yang and others for their contributions to the long-term data collection at IMGERS. This project was supported by the State Key Basic Research and Development Plan of China, the Natural Science Foundation of China and the Knowledge Innovation Program of the Chinese Academy of Sciences.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to J.W. (Jingle.Wu@asu.edu).

## Epigenetic regulation of translation reveals hidden genetic variation to produce complex traits

Heather L. True\*, Ilana Berlin & Susan L. Lindquist

Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142, USA

\* Present address: Washington University Medical School, Department of Cell Biology and Physiology, St Louis, Missouri 63130, USA

Phenotypic plasticity and the exposure of hidden genetic variation both affect the survival and evolution of new traits<sup>1–3</sup>, but their contributing molecular mechanisms are largely unknown. A single factor, the yeast prion  $[PSI^+]$ , may exert a profound effect on both<sup>4</sup>.  $[PSI^+]$  is a conserved, protein-based genetic element that is formed by a change in the conformation and function of the translation termination factor Sup35p<sup>5</sup>, and is transmitted from mother to progeny. Curing cells of  $[PSI^+]$  alters their survival in different growth conditions and produces a spectrum of phenotypes in different genetic backgrounds<sup>4</sup>. Here we show, by examining three plausible explanations for

this phenotypic diversity, that all traits tested involved  $[PSI^+]$ -mediated read-through of nonsense codons. Notably, the phenotypes analysed were genetically complex, and genetic re-assortment frequently converted  $[PSI^+]$ -dependent phenotypes to stable traits that persisted in the absence of  $[PSI^+]$ . Thus,  $[PSI^+]$  provides a temporary survival advantage under diverse conditions, increasing the likelihood that new traits will become fixed by subsequent genetic change. As an epigenetic mechanism that globally affects the relationship between genotype and phenotype,  $[PSI^+]$  expands the conceptual framework for phenotypic plasticity, provides a one-step mechanism for the acquisition of complex traits and affords a route to the genetic assimilation of initially transient epigenetic traits.

There are three possible explanations for the diversity of phenotypes that are produced when different  $[PSI^+]$  strains are cured of the prion by growth on guanidine hydrochloride (GdHCl)<sup>4</sup>. Firstly, several yeast prions have now been identified<sup>5</sup> (and others probably exist)<sup>6,7</sup>, each with different potential biological consequences. As growth on GdHCl cures all known naturally occurring prions in *Saccharomyces cerevisiae*, the diverse phenotypes we observed may have arisen from eliminating different combinations of prions present in different strains. Secondly, the reduction in translation–termination activity that is associated with  $[PSI^+]$  may cause ribosomes to read through naturally occurring stop codons<sup>8–10</sup>. This could alter message stability and/or promote the translation of sequences that are prone to accumulating genetic variation, such as pseudogenes and 3' untranslated regions (see Supplementary Data and Supplementary Fig. 1). Thirdly,  $[PSI^+]$  prion formation is accompanied by protein aggregation. Therefore, the acquisition and removal of such aggregates could have diverse effects on protein homeostasis and produce distinct phenotypes in different strains<sup>11</sup>.

We systematically tested these hypotheses by creating three sets of strain derivatives in several different genetic backgrounds that uncoupled the effects of other prions (set one), translational read-through (set two) and protein aggregation (set three) (Supplementary Table 1). For set one, prions were cured by two different general methods that should eliminate most, if not all, prions (growth on GdHCl and deletion of Hsp104; ref. 5) or by two highly selective methods (mutating the prion-determining domain of the *SUP35* gene (NM) to eliminate  $[PSI^+]$ <sup>12,13</sup> and deleting the *RNQ1* gene to eliminate the other known prion present in some of the strains,  $[RNQ^+]$ <sup>14</sup>). For set two, we recreated the effects of translational read-through in cells that did not contain the prion ( $[psi^-]$ ) either by introducing partial loss-of-function mutations into *SUP35*<sup>15,16</sup> or by introducing mutations that alter the stability of messenger RNAs containing nonsense codons ( $\Delta upf1$ ,  $\Delta ski7$ )<sup>17,18</sup>. For set three, prion-like aggregates were recreated in a strain immune to their effects on translation termination. The strains were selectively cured of  $[PSI^+]$  by chromosomal deletion of the prion-forming domain (*sup35* <sup>$\Delta$ NM</sup>) and NM–green fluorescent protein (GFP) was expressed from a strong promoter. We also created other strains that retained Sup35p prion aggregates but had efficient translation termination. This was accomplished by introducing a form of Sup35p that is immune to capture by  $[PSI^+]$  aggregates (extra-chromosomal expression of *sup35C* (without NM)<sup>12</sup> or antisuppressor (ASU) *sup35* variants)<sup>19</sup>.

The majority of the phenotypes tested proved to be a simple and direct consequence of  $[PSI^+]$ -mediated nonsense suppression (Table 1; Fig. 1; Supplementary Fig. 2; and data not shown). For example, resistance to 3 mM paraquat was greater in the  $[PSI^+]$  derivative of strain 5V-H19 than in the  $[psi^-]$  derivative (Fig. 1, centre). The  $[psi^-]$  cells acquired resistance to paraquat in the absence of the prion when wild-type *SUP35* was replaced with a mutant that enhanced nonsense suppression (*sup35*<sup>C653R</sup>).  $[PSI^+]$  cells lost paraquat resistance when  $[PSI^+]$  was selectively cured by deleting the prion-forming domain ( $\Delta$ NM). Reducing nonsense

suppression with the expression of either *Sup35C* or the anti-suppressor (ASU) mutant *sup35<sup>Q15R</sup>* in the original [*PSI<sup>+</sup>*] derivative masked the phenotype. However, re-introducing NM aggregates in the [*psi<sup>-</sup>*] derivative did not restore paraquat resistance (*sup35<sup>ΔNM</sup>* with the NM-GFP construct, data not shown).

Similar observations were made in cases where [*PSI<sup>+</sup>*] conferred a growth disadvantage. The 5V-H19 [*PSI<sup>+</sup>*] derivative displayed a greater sensitivity to 100 mM hydroxyurea (HU) than the [*psi<sup>-</sup>*] derivative (Fig. 1, right). Simply increasing nonsense suppression in the [*psi<sup>-</sup>*] derivative (with *sup35<sup>C653R</sup>*) recapitulated HU sensitivity. Either eliminating [*PSI<sup>+</sup>*] selectively (*ΔNM*) or decreasing nonsense suppression in a [*PSI<sup>+</sup>*] strain (with extra-chromosomal *sup35C* or *sup35<sup>Q15R</sup>*) restored HU resistance to levels similar to those of the [*psi<sup>-</sup>*] derivative.

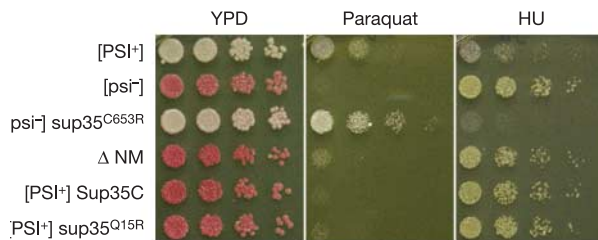
Some phenotypes were dependent on read-through, but could not be fully attributed to this (Table 1). For example, the phenotypes were lost when cells were selectively cured of [*PSI<sup>+</sup>*] or when translation termination efficiency was increased in [*PSI<sup>+</sup>*], but they were not restored in [*psi<sup>-</sup>*] cells simply by increasing translational read-through. In order to increase read-through in [*psi<sup>-</sup>*] strains we used different *sup35* mutations<sup>15,16</sup> to approximate the different levels of read-through that characterized the original [*PSI<sup>+</sup>*] derivative in each genetic background (verified by nonsense suppression assays<sup>20</sup>; Supplementary Table 2; and data not shown)<sup>4</sup>. Either more precise levels of read-through are required to reproduce some phenotypes, or additional factors (such as other prions or the presence of aggregates) contribute to these traits.

Notably, neither other prions nor aggregation alone could account for any of the phenotypes tested. The selective curing of [*PSI<sup>+</sup>*] eliminated each phenotype as effectively as curing by general methods. Selectivity in the curing of [*PSI<sup>+</sup>*] was verified by the retention of [*RNQ<sup>+</sup>*] in strains known to carry it (Supplementary Table 1; data not shown). Moreover, re-establishing NM-containing aggregates by overexpressing NM-GFP in cells cured of [*PSI<sup>+</sup>*] (*sup35<sup>ΔNM</sup>*) did not restore any of the phenotypes. Finally, we analysed diploids created by outcrosses between the [*PSI<sup>+</sup>*] strains of various genetic backgrounds and found that the phenotypes were recessive in nature. If they had been a direct consequence of the presence of different prions or protein aggregates in different genetic backgrounds they would be predicted to be dominant (Supplementary Table 3; data not shown). Taken together, the data strongly suggest that most [*PSI<sup>+</sup>*]-associated phenotypes are due to the interaction of particular read-through events with the genetic architecture of the original strain, rather

than the presence of other prions or protein aggregation *per se*.

Our data suggest that all phenotypes are associated with translational read-through, but they may be acquired through a variety of molecular mechanisms. A probable source of the phenotypic diversity is the expression of information in previously silent regions of the genome, which are free to acquire genetic variation such as pseudogenes and 3' sequences at the ends of normal mRNAs (Table 1; Supplementary Data and Supplementary Fig. 1). Translational read-through by [*PSI<sup>+</sup>*] may also affect other pathways, such as the system that turns over mRNAs when ribosomes read through the last stop codon at the end of a message (the non-stop decay pathway<sup>18</sup> through *SKI7*), or the system that turns over mRNAs when ribosomes stop prematurely (the nonsense mediated decay pathway<sup>17</sup> through *UPF1*). Indeed, some traits (but not all) were recapitulated by the deletion of *UPF1* in [*psi<sup>-</sup>*] cells, which stabilizes transcripts containing premature stop codons (Table 1, *Δupf1*).

To investigate the genetic complexity of [*PSI<sup>+</sup>*]-dependent phenotypes, strains were outcrossed to genetic backgrounds that did not have the same [*PSI<sup>+</sup>*]-dependent phenotype, and meiotic progeny were analysed. Among the many phenotypes assessed, we did not find a single case of simple 2:2 segregation of the trait (Supplementary Table 3; and data not shown). Moreover, the progeny invariably presented a broad range of phenotypes of variable strengths, including some with stronger phenotypes than the parental strains (Fig. 2; Supplementary Table 3). For example, to analyse the resistance to 3 mM paraquat of the [*PSI<sup>+</sup>*] derivative of the strain 5V-H19, we outcrossed it to D1142-1A. D1142-1A does



**Figure 1** [*PSI<sup>+</sup>*]-related phenotypes are dependent on read-through. Nonsense suppression enhances resistance to 3 mM paraquat and increases sensitivity to 100 mM hydroxyurea (HU) in the strain 5V-H19 (in [*PSI<sup>+</sup>*] and [*psi<sup>-</sup>*] *sup35<sup>C653R</sup>*). Elimination of nonsense suppression (in [*psi<sup>-</sup>*], *ΔNM*, [*PSI<sup>+</sup>*] *Sup35C* and [*PSI<sup>+</sup>*] *sup35<sup>Q15R</sup>*) results in sensitivity to paraquat and increased resistance to HU.

**Table 1** [*PSI<sup>+</sup>*]-dependent phenotypes across genetic backgrounds

| Phenotype                                                           | Strain background | Resistance                 | Sensitivity                | Suppression method(s)                                                         | Anti-suppression method(s)                                             |
|---------------------------------------------------------------------|-------------------|----------------------------|----------------------------|-------------------------------------------------------------------------------|------------------------------------------------------------------------|
| Phenotypes solely dependent on read-through                         |                   |                            |                            |                                                                               |                                                                        |
| Caffeine 10 mM                                                      | 5V-H19            | [ <i>PSI<sup>+</sup></i> ] | [ <i>psi<sup>-</sup></i> ] | [ <i>psi<sup>-</sup></i> ] <i>sup35<sup>C653R</sup></i> , <i>Δupf1</i>        | [ <i>PSI<sup>+</sup></i> ] <i>Sup35C</i> , <i>Sup35<sup>Q15R</sup></i> |
| Caffeine 10 mM                                                      | 74-D694           | [ <i>PSI<sup>+</sup></i> ] | [ <i>psi<sup>-</sup></i> ] | [ <i>psi<sup>-</sup></i> ] <i>sup35<sup>R8</sup></i>                          | [ <i>PSI<sup>+</sup></i> ] <i>Sup35C</i> , <i>Sup35<sup>Q15R</sup></i> |
| Caffeine 5 mM                                                       | SL1010-1A         | [ <i>PSI<sup>+</sup></i> ] | [ <i>psi<sup>-</sup></i> ] | <i>Δupf1</i> [ <i>psi<sup>-</sup></i> ]                                       | [ <i>PSI<sup>+</sup></i> ] <i>Sup35C</i> , <i>Sup35<sup>Q15R</sup></i> |
| Paraquat 3 mM                                                       | 5V-H19            | [ <i>PSI<sup>+</sup></i> ] | [ <i>psi<sup>-</sup></i> ] | [ <i>psi<sup>-</sup></i> ] <i>sup35<sup>C653R</sup></i> , <i>Δupf1</i>        | [ <i>PSI<sup>+</sup></i> ] <i>Sup35C</i> , <i>Sup35<sup>Q15R</sup></i> |
| Paraquat 4 mM                                                       | Bsc 783/4c        | [ <i>PSI<sup>+</sup></i> ] | [ <i>psi<sup>-</sup></i> ] | [ <i>psi<sup>-</sup></i> ] <i>sup35<sup>R320I</sup></i>                       | [ <i>PSI<sup>+</sup></i> ] <i>Sup35C</i> , <i>Sup35<sup>Q15R</sup></i> |
| LiCl 100 mM                                                         | 5V-H19            | [ <i>PSI<sup>+</sup></i> ] | [ <i>psi<sup>-</sup></i> ] | [ <i>psi<sup>-</sup></i> ] <i>sup35<sup>C653R</sup></i>                       | [ <i>PSI<sup>+</sup></i> ] <i>Sup35C</i> , <i>Sup35<sup>Q15R</sup></i> |
| Benomyl 1 μg ml <sup>-1</sup>                                       | 10B-H49           | [ <i>PSI<sup>+</sup></i> ] | [ <i>psi<sup>-</sup></i> ] | <i>Δupf1</i> [ <i>psi<sup>-</sup></i> ]                                       | N/A                                                                    |
| Ethanol 10%                                                         | 5V-H19            | [ <i>PSI<sup>+</sup></i> ] | [ <i>psi<sup>-</sup></i> ] | [ <i>psi<sup>-</sup></i> ] <i>sup35<sup>C653R</sup></i> , <i>Δupf1</i>        | [ <i>PSI<sup>+</sup></i> ] <i>Sup35C</i> , <i>Sup35<sup>Q15R</sup></i> |
| Ethanol 10%                                                         | D1142-1A          | [ <i>PSI<sup>+</sup></i> ] | [ <i>psi<sup>-</sup></i> ] | <i>Δupf1</i> [ <i>psi<sup>-</sup></i> ]                                       | N/A                                                                    |
| Hydroxyurea 100 mM                                                  | 5V-H19            | [ <i>psi<sup>-</sup></i> ] | [ <i>PSI<sup>+</sup></i> ] | [ <i>psi<sup>-</sup></i> ] <i>sup35<sup>C653R</sup></i>                       | [ <i>PSI<sup>+</sup></i> ] <i>Sup35C</i> , <i>Sup35<sup>Q15R</sup></i> |
| ZnCl <sub>2</sub> 7.5 mM                                            | 74-D694           | [ <i>psi<sup>-</sup></i> ] | [ <i>PSI<sup>+</sup></i> ] | [ <i>psi<sup>-</sup></i> ] <i>sup35<sup>R8</sup></i>                          | [ <i>PSI<sup>+</sup></i> ] <i>Sup35C</i> , <i>Sup35<sup>Q15R</sup></i> |
| LiCl 400 mM                                                         | 74-D694           | [ <i>psi<sup>-</sup></i> ] | [ <i>PSI<sup>+</sup></i> ] | [ <i>psi<sup>-</sup></i> ] <i>sup35<sup>R8</sup></i>                          | [ <i>PSI<sup>+</sup></i> ] <i>Sup35C</i> , <i>Sup35<sup>Q15R</sup></i> |
| Phenotypes dependent on read-through but may involve other factors* |                   |                            |                            |                                                                               |                                                                        |
| CsCl 150 mM                                                         | Bsc 783/4c        | [ <i>PSI<sup>+</sup></i> ] | [ <i>psi<sup>-</sup></i> ] | N/O ([ <i>psi<sup>-</sup></i> ] <i>sup35<sup>R320I</sup></i> , <i>Δupf1</i> ) | [ <i>PSI<sup>+</sup></i> ] <i>Sup35C</i> , <i>Sup35<sup>Q15R</sup></i> |
| CdCl <sub>2</sub> 75 μM                                             | 5V-H19            | [ <i>psi<sup>-</sup></i> ] | [ <i>PSI<sup>+</sup></i> ] | [ <i>psi<sup>-</sup></i> ] <i>sup35<sup>C653R</sup></i> , <i>Δupf1</i>        | P ([ <i>PSI<sup>+</sup></i> ] <i>Sup35<sup>Q15R</sup></i> )            |
| Paromomycin 200 μg ml <sup>-1</sup>                                 | 5V-H19            | [ <i>psi<sup>-</sup></i> ] | [ <i>PSI<sup>+</sup></i> ] | N/O ([ <i>psi<sup>-</sup></i> ] <i>sup35<sup>C653R</sup></i> , <i>Δupf1</i> ) | [ <i>PSI<sup>+</sup></i> ] <i>Sup35C</i> , <i>Sup35<sup>Q15R</sup></i> |
| ZnCl <sub>2</sub> 7.5 mM                                            | SL1010-1A         | [ <i>psi<sup>-</sup></i> ] | [ <i>PSI<sup>+</sup></i> ] | N/O ( <i>Δupf1</i> [ <i>psi<sup>-</sup></i> ])                                | [ <i>PSI<sup>+</sup></i> ] <i>Sup35C</i> , <i>Sup35<sup>Q15R</sup></i> |
| ZnCl <sub>2</sub> 5 mM                                              | 5V-H19            | [ <i>psi<sup>-</sup></i> ] | [ <i>PSI<sup>+</sup></i> ] | N/O ([ <i>psi<sup>-</sup></i> ] <i>sup35<sup>C653R</sup></i> , <i>Δupf1</i> ) | [ <i>PSI<sup>+</sup></i> ] <i>Sup35C</i> , <i>Sup35<sup>Q15R</sup></i> |

\* Phenotypes partially effected in suppression strains and/or not able to recapitulate with nonsense suppression alone denoted in parentheses.

N/A: not available in the given strain background.

N/O: intended suppression effect is not observed in respect to the given phenotype.

P: partial effect observed.

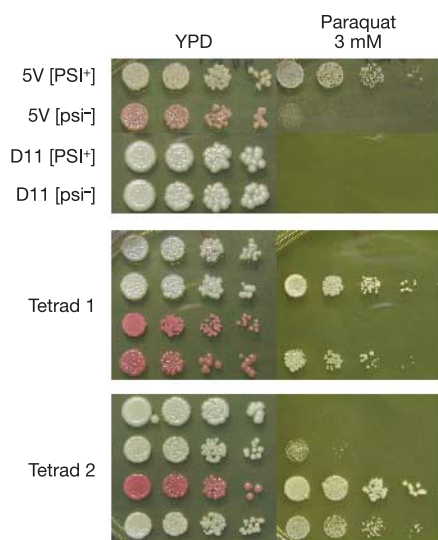


not grow on 3 mM paraquat in either the  $[PSI^+]$  or the  $[psi^-]$  state. Segregation of paraquat resistance was 2:2 for only two of the sixteen tetrads analysed (Supplementary Table 3) and progeny in many tetrads displayed intermediate phenotypes (Fig. 2; Supplementary Table 3). Similar analyses were completed for the  $[PSI^+]$ -dependent caffeine-resistant traits of both 74-D694 and 5V-H19 (Supplementary Table 3). Each strain was outcrossed to Bsc783/4c and W303-1A, neither of which is able to grow in the presence of 10 mM caffeine. None of the four outcrosses yielded a simple 2:2 pattern of segregation for caffeine resistance and many progeny of intermediate phenotypes were obtained. These and other data (Supplementary Table 3; data not shown) suggest that the  $[PSI^+]$ -dependent phenotypes can be complex traits acquired through a combination of genetic changes (some of which are hidden but are uncovered by translational read through events). Thus,  $[PSI^+]$  provides a route to the acquisition of complex traits in a single step.

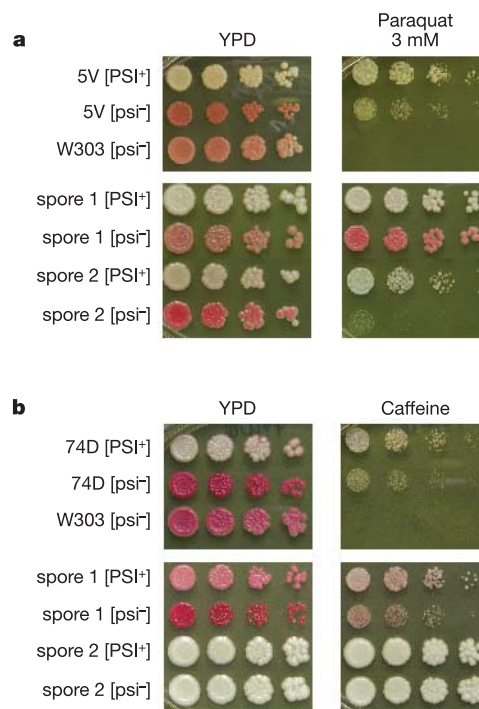
Both the acquisition of these heritable new states (when cells become  $[PSI^+]$ ) and the loss of those states (when cells revert to the  $[psi^-]$  state) are accomplished without changes in the genome, providing a transient mechanism for survival in fluctuating environments (Supplementary Fig. 3). However, the potential impact of the genetic variation accessed in the  $[PSI^+]$  state would be much greater if it were susceptible to fixation and maintenance in the absence of the prion. Indeed, further analysis of the outcross progeny showed that fixation of the complex  $[PSI^+]$ -mediated traits occurs readily. In fact, it can occur as a result of a single cross. The original phenotypes we examined were dependent upon  $[PSI^+]$ , but in many outcross progeny the phenotypes were less dependent on (or even independent of) the prion. For example, the phenotype was maintained after cells were cured of  $[PSI^+]$  by either growth on GdHCl or by deletion of *HSP104*. Also, after 5V-H19  $[PSI^+]$  (Fig. 3a) and 74-D694  $[PSI^+]$  (Fig. 3b) were outcrossed to the caffeine sensitive (caffeine<sup>s</sup>) strain W303-1A, many progeny maintained caffeine resistance after curing. In each of the phenotypes analysed to this extent (Supplementary Table 3), approximately 50% of the progeny exhibiting the trait maintained it in a  $[PSI^+]$ -dependent manner, whereas the other 50% maintained it in the absence of  $[PSI^+]$ . That is, the trait was fixed. Subsequent

analysis demonstrated that there were multiple mechanisms for fixation of the same trait. Moreover, fixation was not due to the simple appearance of global nonsense suppressors (Supplementary Data). Rather, the phenotypes were fixed by virtue of the re-assortment of other genetic polymorphisms, by the appearance of new mutations or by a combination of the two.

Conceptually,  $[PSI^+]$  presents a new framework for phenotypic plasticity. Because  $[PSI^+]$  is a metastable element that is both gained and lost at a low spontaneous rate ( $10^{-5}$  to  $10^{-7}$ ), large populations of yeast are expected to contain both  $[PSI^+]$  and  $[psi^-]$  derivatives<sup>21</sup>. Individuals at a local peak of an adaptive landscape can suddenly cross to a new adaptive peak and survive when the environment changes. This change in state may be influenced by environmental conditions<sup>22</sup>, but also occurs in a stochastic fashion through a self-perpetuating change in the folding of a single protein. Reduced translational efficiency in general, as well as many of the individual genetic variants it uncovers is likely to be detrimental (see Supplementary Data), but the capacity to acquire the  $[PSI^+]$  state has been conserved in yeast for at least a hundred million years<sup>23–26</sup>. The rarity of the change in state<sup>21</sup> ensures that the detrimental impact of the lower translational fidelity is negligible with respect to the survival of the population as a whole<sup>27</sup>. However, the ability to switch to the  $[PSI^+]$  state enables cells to benefit from a hidden potential to adapt and survive in particular circumstances. This increases the population size of the organisms with that genome and in turn, increases the chances for assimilation of the trait. Assimilation could occur by genetic re-assortment through mating (as shown here), by the mutation of stop to sense codons or by the acquisition of mutations that alter the stability of mRNAs (Supplementary Data and Supplementary Fig. 3).



**Figure 2** Variable strength phenotypes are observed with progeny from outcrosses. Shown here is the variability of resistance in two tetrads from the cross of 5V-H19  $[PSI^+]$  to D1142-1A  $[PSI^+]$  on YPD with 3 mM paraquat. Note that suppression of *ade2-1* by  $[PSI^+]$ , in combination with the suppressor transfer RNA *SUQ5*, results in an alteration of pigment of colonies grown on standard media, such that  $[PSI^+]$  colonies are white with the combination of these markers. Therefore, the absence of co-segregation of *ade2-1* with *SUQ5* results in some red  $[PSI^+]$  colonies in outcross progeny.



**Figure 3** Progeny from an outcross of 5V-H19 show  $[PSI^+]$ -independent caffeine resistance. **a**,  $[PSI^+]$ -dependent resistance to 10 mM caffeine in YPD in the parental strain 5V-H19 and caffeine sensitivity in W303-1A is shown. Some progeny were resistant to caffeine irrespective of the  $[PSI^+]$  state, whereas other progeny maintained  $[PSI^+]$ -dependent resistance. The  $[psi^-]$  variants of spores were derived by either guanidine hydrochloride curing of  $[PSI^+]$  spores, or by deletion of *HSP104*. **b**,  $[PSI^+]$ -dependent resistance to 10 mM caffeine in YPD in the parental strain 74-D694 and caffeine sensitivity in W303-1A is shown along with two progeny in both the  $[PSI^+]$  and  $[psi^-]$  states.

For organisms to survive in the short term and evolve in the long term they must endure fluctuating environments, and then assimilate genetic changes that produce new traits. Must they wait for new mutations to occur in order to produce new traits? Or are there mechanisms that allow organisms to draw on pre-existing genetic variation? We have previously shown that environmental stress can reveal hidden genetic variation in a global genome-wide fashion by altering the activities of a protein chaperone, Hsp90<sup>28,29</sup>, which functions in multiple signal transduction pathways. Here we demonstrate that a very different mechanism, a self-perpetuating change in the function and conformation of Sup35p, can also reveal previously hidden genetic variation in a combinatorial fashion on a global scale. In contrast to Hsp90, it has the capacity to do so in a single step that is immediately heritable but not yet fixed. Taking these observations together, we suggest that protein folding mechanisms are likely to have important and unexpected roles in translating genotypes to phenotypes and are likely to influence evolution in ways that could not have been foreseen. □

## Methods

### Strain construction

All yeast strains used in this study are given in Supplementary Table 1. All yeast genetic and molecular biology manipulations were done as previously described<sup>30</sup>. In the [*PSI*<sup>+</sup>] state, translation termination efficiency is reduced and the loss of function can be detected by the suppression of nonsense codon mutations in selectable markers<sup>5</sup>. All original [*psi*<sup>-</sup>] variants were obtained from their isogenic [*PSI*<sup>+</sup>] counterparts by curing on 3 mM GdHCl in complete rich medium (YPD).  $\Delta$ rnq1 mutants were generated independently in both [*PSI*<sup>+</sup>] and [*psi*<sup>-</sup>] backgrounds,  $\Delta$ upf1 and  $\Delta$ ski7 in [*psi*<sup>-</sup>] and  $\Delta$ hsp104 in [*PSI*<sup>+</sup>] according to standard protocols. All open reading frame (ORF) disruptions were verified by polymerase chain reaction (PCR). The  $\Delta$ rnq1 and  $\Delta$ hsp104 strains tested negative for Rnq1p and Hsp104p respectively, as assayed by western blot analysis. Conservation of [RNQ<sup>+</sup>] prion was verified by solubility assays as described<sup>14</sup>.

To selectively cure [*PSI*<sup>+</sup>], deletion mutants were generated in [*PSI*<sup>+</sup>] backgrounds (*sup35*<sup>ΔNM</sup> and *sup35*<sup>Δ2-5</sup>; Supplementary Table 1) by a replacement of wild type *SUP35* with pSup35C and pSup35RΔ2-5 (ref. 13) respectively using a pop-in/pop-out method. Both sets of mutants were verified by PCR, and *sup35*<sup>ΔNM</sup> strains tested negative for the presence of the NM region of Sup35p by western blot analysis. The pTEF-NM-GFP construct was integrated into *sup35*<sup>ΔNM</sup> background at *URA3*, creating ΔNM/NM-GFP. The presence of NM-GFP aggregates was confirmed *in vivo* by fluorescence microscopy.

The *sup35* mutants (C653R (ref. 16), R320I (ref. 16) and *sup35*-R8 (ref. 15)) were introduced into [*psi*<sup>-</sup>] strains by replacement of the wild type gene. Exogenous *sup35C* was introduced by integration of pTEF-sup35C into the genome of [*PSI*<sup>+</sup>] variants at *URA3*. The anti-suppressor mutants *sup35*<sup>Q15R</sup> and *sup35*<sup>S17R</sup> were integrated into the [*PSI*<sup>+</sup>] strains at *LEU2* using pASU<sup>Q15R</sup> and pASU<sup>S17R</sup>, respectively. The presence of [*PSI*<sup>+</sup>]-like aggregates in the anti-suppressor strains was verified by *in vivo* aggregates observed with overexpression of NM-GFP. The suppression levels of the strains described above were assayed using the read through assay<sup>20</sup> and by growth on solid minimal media (SD) lacking adenine. The data shown in Supplementary Tables 2 and 4 are an average of two independent transformants grown and assayed on the same day. The absolute numbers varied from day to day, but the trend was always reproducible.

### DNA preparation

DNA cloning and PCR were performed using standard methods. Oligonucleotides were obtained from Operon. Sequencing was done by Northwoods DNA Sequencing.

### Plasmid construction

The *sup35* mutants (C653R, R320I and *sup35*-R8) were cloned from previously characterized strains<sup>15,16</sup> and inserted into pRS306. pSup35C was constructed by PCR amplification of the carboxy-terminal region of *SUP35*<sup>12</sup> as a *Bam*HI/*Sac*I fragment and inserted into a pRS306 cassette containing the endogenous *SUP35* promoter at *Eco*RI/*Bam*HI. PCR primers used were as follows: 5'-CCGGAATTCATGTCGG ATTCAAACCAAGGC-3'; 3'-Sup35XhoI: 5'-CCGCTCGAGTTACTCGCAATTTTAA CAATTTTACC-3'; 3'-GFPXhoI: 5'-CCGCTCGAGTCATTGTATAGTTTCATCAATG-3'. The plasmid pTEF-Sup35C was constructed by substituting the *SUP35* promoter of pSup35C with the TEF promoter (*Sac*I/*Bam*HI) from p416TEF at PmeI/*Bam*HI by blunt-end ligation. To construct pNMG, NM-GFP was amplified by PCR as an *Eco*RI-XhoI fragment and inserted into a pRS306-TEF vector containing the promoter from p416TEF at *Sac*I and *Xba*I. Anti-suppressors pASUQ15R and pASUS17R were obtained by cloning Sup35Pr-ASU-EF from original vectors<sup>19</sup> as XhoI-SacI fragments into a pRS306 vector.

### Phenotypic testing

To test strain sets under various stress conditions, cells were grown to late log phase in YPD and fivefold serial dilutions were spotted on solid media containing appropriate treatments, using a replicator as described previously<sup>4</sup>. The phenotypes shown are reproducible using specific media in the concentrations indicated for various compounds. Alterations in strength and appearance of phenotype are typical with different media and different concentrations of the same additives.

Received 13 April; accepted 23 July 2004; doi:10.1038/nature02885.

Published online 15 August 2004.

- Agrawal, A. A. Phenotypic plasticity in the interactions and evolution of species. *Science* **294**, 321–326 (2001).
- Behera, N. & Nanjundiah, V. Phenotypic plasticity can potentiate rapid evolutionary change. *J. Theor. Biol.* **226**, 177–184 (2004).
- Pigliucci, M. & Murren, C. J. Perspective: Genetic assimilation and a possible evolutionary paradox: can macroevolution sometimes be so fast as to pass us by? *Evolution Int. J. Org. Evolution* **57**, 1455–1464 (2003).
- True, H. L. & Lindquist, S. L. A yeast prion provides a mechanism for genetic variation and phenotypic diversity. *Nature* **407**, 477–483 (2000).
- Uptain, S. M. & Lindquist, S. Prions as protein-based genetic elements. *Annu. Rev. Microbiol.* **56**, 703–741 (2002).
- Michelitsch, M. D. & Weissman, J. S. A census of glutamine/asparagine-rich regions: implications for their conserved function and the prediction of novel prions. *Proc. Natl Acad. Sci. USA* **97**, 11910–11915 (2000).
- Harrison, P. M. & Gerstein, M. A method to assess compositional bias in biological sequences and its application to prion-like glutamine/asparagine-rich domains in eukaryotic proteomes. *Genome Biol.* **4**, R40 (2003).
- Harrison, P. et al. A small reservoir of disabled ORFs in the yeast genome and its implications for the dynamics of proteome evolution. *J. Mol. Biol.* **316**, 409–419 (2002).
- Namy, O., Duchateau-Nguyen, G. & Rousset, J. P. Translational readthrough of the PDE2 stop codon modulates cAMP levels in *Saccharomyces cerevisiae*. *Mol. Microbiol.* **43**, 641–652 (2002).
- Namy, O. et al. Identification of stop codon readthrough genes in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **31**, 2289–2296 (2003).
- Sherman, M. Y. & Goldberg, A. L. Cellular defenses against unfolded proteins: a cell biologist thinks about neurodegenerative diseases. *Neuron* **29**, 15–32 (2001).
- Ter-Avanesyan, M. D. et al. Deletion analysis of the SUP35 gene of the yeast *Saccharomyces cerevisiae* reveals two non-overlapping functional regions in the encoded protein. *Mol. Microbiol.* **7**, 683–692 (1993).
- Liu, J. J. & Lindquist, S. Oligopeptide-repeat expansions modulate 'protein-only' inheritance in yeast. *Nature* **400**, 573–576 (1999).
- Sondheimer, N. & Lindquist, S. Rnq1: an epigenetic modifier of protein function in yeast. *Mol. Cell* **5**, 163–172 (2000).
- Chernoff, Y. O., Newnam, G. P. & Liebman, S. W. The translational function of nucleotide C1054 in the small subunit rRNA is conserved throughout evolution: genetic evidence in yeast. *Proc. Natl Acad. Sci. USA* **93**, 2517–2522 (1996).
- Bradley, M. E., Bagriantsev, S., Vishveshwara, N. & Liebman, S. W. Guanidine reduces stop codon read-through caused by missense mutations in SUP35 or SUP45. *Yeast* **20**, 625–632 (2003).
- Leeds, P., Peltz, S. W., Jacobson, A. & Culbertson, M. R. The product of the yeast UPF1 gene is required for rapid turnover of mRNAs containing a premature translational termination codon. *Genes Dev.* **5**, 2303–2314 (1991).
- Frischmeyer, P. A. et al. An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. *Science* **295**, 2258–2261 (2002).
- DePace, A. H., Santoso, A., Hillner, P. & Weissman, J. S. A critical role for amino-terminal glutamine/asparagine repeats in the formation and propagation of a yeast prion. *Cell* **93**, 1241–1252 (1998).
- Firoozan, M., Grant, C. M., Duarte, J. A. & Tuite, M. F. Quantitation of readthrough of termination codons in yeast using a novel gene fusion assay. *Yeast* **7**, 173–183 (1991).
- Lund, P. M. & Cox, B. S. Reversion analysis of [*psi*<sup>-</sup>] mutations in *Saccharomyces cerevisiae*. *Genet. Res.* **37**, 173–182 (1981).
- Eaglestone, S. S., Cox, B. S. & Tuite, M. F. Translation termination efficiency can be regulated in *Saccharomyces cerevisiae* by environmental stress through a prion-mediated mechanism. *EMBO J.* **18**, 1974–1981 (1999).
- Chernoff, Y. O. et al. Evolutionary conservation of prion-forming abilities of the yeast Sup35 protein. *Mol. Microbiol.* **35**, 865–876 (2000).
- Kushnirov, V. V., Kochneva-Pervukhova, N. V., Chechenova, M. B., Frolova, N. S. & Ter-Avanesyan, M. D. Prion properties of the Sup35 protein of yeast *Pichia methanolicola*. *EMBO J.* **19**, 324–331 (2000).
- Nakayashiki, T., Ebihara, K., Bannai, H. & Nakamura, Y. Yeast [*PSI*<sup>+</sup>] "prions" that are cross-transmissible and susceptible beyond a species barrier through a quasi-prion state. *Mol. Cell* **7**, 1121–1130 (2001).
- Santoso, A., Chien, P., Osherovich, L. Z. & Weissman, J. S. Molecular basis of a yeast prion species barrier. *Cell* **100**, 277–288 (2000).
- Masel, J. & Bergman, A. The evolution of the evolvability properties of the yeast prion [*PSI*<sup>+</sup>]. *Evolution* **57**, 1498–1512 (2003).
- Rutherford, S. L. & Lindquist, S. Hsp90 as a capacitor for morphological evolution. *Nature* **396**, 336–342 (1998).
- Queitsch, C., Sangster, T. A. & Lindquist, S. Hsp90 as a capacitor of phenotypic variation. *Nature* **417**, 618–624 (2002).
- Guthrie, C. & Fink, G. (eds) *Guide to Yeast Genetics and Molecular Biology* (Academic Press, San Diego, 1991).

**Supplementary Information** accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank members of the Lindquist lab, the True lab, R. Parker and J. Masel for discussions and comments on experiments and the manuscript. We thank J. Weissman, S. Liebman and their lab members for plasmids and strains. This research was supported by the National Institutes of Health.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to S.L.L. (lindquist\_admin@wi.mit.edu).

# Recollection-like memory retrieval in rats is dependent on the hippocampus

Norbert J. Fortin, Sean P. Wright & Howard Eichenbaum

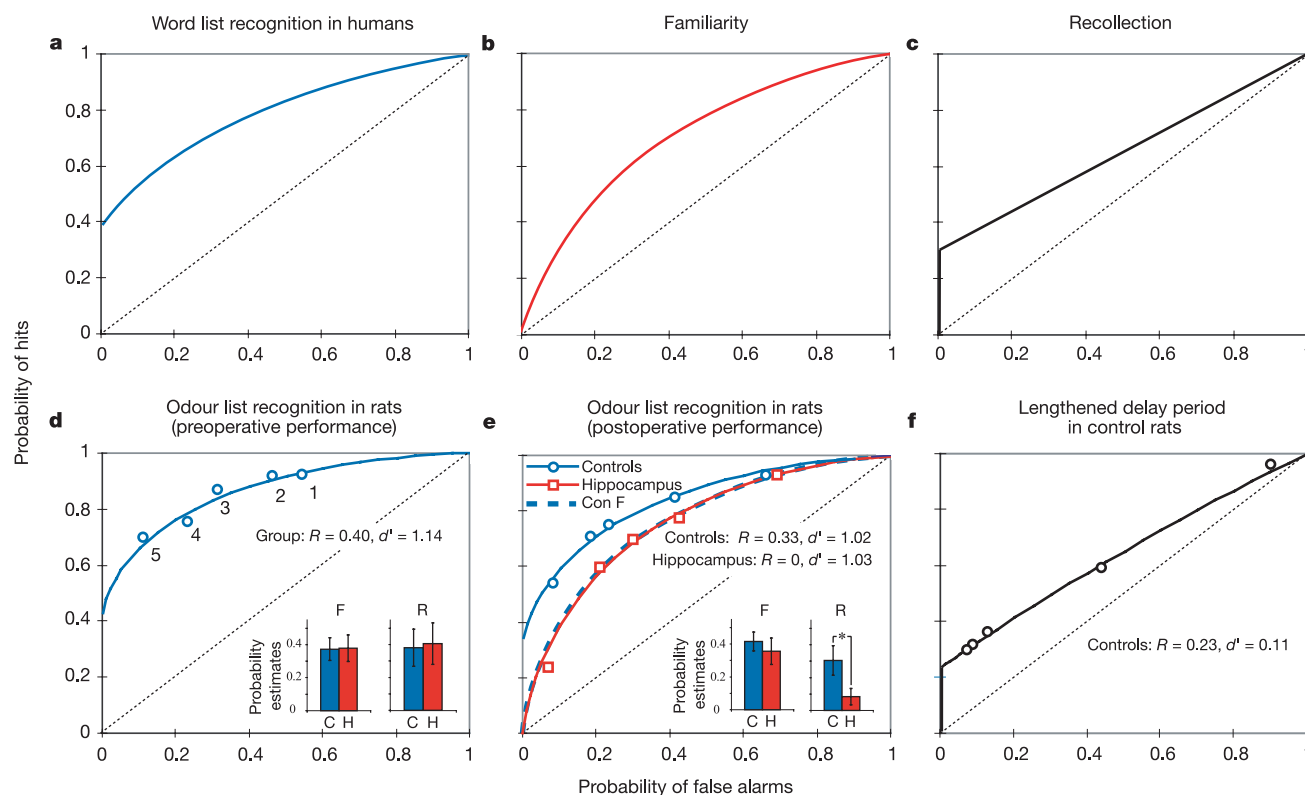
Center for Memory and Brain, Boston University, 2 Cummington Street, Boston, Massachusetts 02215, USA

Recognition memory may be supported by two independent types of retrieval, conscious recollection of a specific experience and a sense of familiarity gained from previous exposure to particular stimuli<sup>1,2</sup>. In humans, signal detection techniques have been used to distinguish recollection and familiarity, respectively, in asymmetrical and curvilinear components of their receiver operating characteristic (ROC) curves, standard curves that represent item recognition across different levels of confidence or bias. To determine whether animals also employ multiple processes in recognition memory and to explore the anatomical basis of this distinction, we adapted these techniques to examine odour recognition memory in rats. Their ROC curve had asymmetrical and curvilinear components, indicating the existence of both recollection and familiarity in rats. Furthermore, following selective damage to the hippocampus the ROC curve became entirely symmetrical and remained curvilinear, supporting the view that the hippocampus specifically mediates the capacity for recollection.

When meeting people on a second occasion, we sometimes recall

the previous encounter fully but at other times experience only a sense of familiarity without recollective experience. Studies on human amnesic patients and functional neuroimaging in normal human subjects have suggested that the hippocampus may be critical to recollection whereas familiarity may be mediated by the surrounding cortical areas<sup>3–5</sup>. However, this proposal is controversial<sup>6,7</sup>, and definitive evidence on the specific role of the hippocampus is beyond the anatomical resolution of current neuropsychological and functional imaging studies on humans. Animal models are well suited for investigations of the specific contribution of the hippocampus in recognition memory because they allow circumscribed brain damage to be generated experimentally, and previous studies on animals have indicated functional distinctions between the hippocampus and adjacent cortical areas<sup>8,9</sup>. However, studies on animals have also failed to provide consensus in that the severity and rate of forgetting in recognition deficits following hippocampal damage is highly variable across different tests<sup>10–17</sup>. A possible explanation is that the hippocampus mediates only the recollective component of recognition and any residual accuracy in performance is due to extra-hippocampal familiarity processes.

Our strategy in testing this hypothesis takes advantage of differences in the retrieval dynamics of recollection and familiarity. In several theoretical models, familiarity is indexed by graded memory strength supporting a precise match between a current stimulus and stored memory, whereas recollection involves a threshold for retrieval of associative and contextual information about an event (for review see ref. 1). An approach that has been particularly useful in distinguishing recollection and familiarity on the basis of these differences is the application of signal detection theory in the analysis of ROCs. In these studies, subjects are initially presented



**Figure 1** ROCs for recognition performance in humans and rats. **a–c**, Performance of humans in verbal recognition (adapted from ref. 20). **d–f**, Performance of rats in odour recognition. **d**, Normal rats tested with a 30-min delay. Insets: recollection (R) and familiarity (F) estimates. **e**, Postoperative performance with a 30-min delay, including an

estimated curve for controls based on familiarity alone (con F). **f**, Control rats tested with a 75-min memory delay. Diagonal dotted lines represent chance performance across criterion levels. C, control group; H, hippocampal group. Error bars,  $\pm$ s.e.m.; asterisk,  $P < 0.05$ .



with a list of items to remember, then following a delay, are presented with the same (old) items as well as new items. Recognition performance is scored in terms of hits (correct identification of old items) and false alarms (misidentification of new items, as though they were old). ROC curves relate the proportion of hits to false alarms across a range of confidence levels or response criteria<sup>18</sup>. In the latter case, the curve varies from a 'liberal' response criterion, where subjects make a high proportion of hits but also generate many false alarms (Fig. 1a, upper right corner), to a 'conservative' criterion, where subjects generate few false alarms but also make a smaller proportion of hits (lower left corner). Chance performance at intermediate criteria falls along the diagonal whereas successful recognition results in an ROC function above the chance line, reflecting more hits than false alarms.

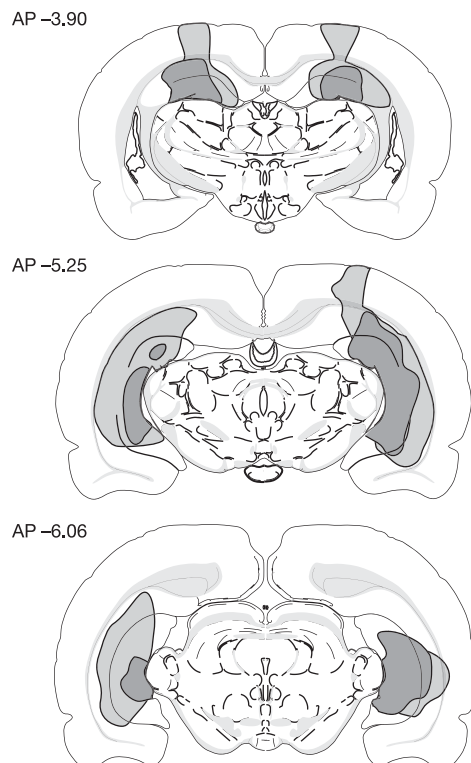
Supporting the theoretical distinctions in retrieval dynamics, ROC analyses have revealed an overall asymmetric recognition performance pattern (Fig. 1a), composed of two independent processes: a symmetrical, curvilinear component associated with familiarity (Fig. 1b) and an asymmetrical, linear component marked by above-zero threshold recollection even under the most stringent bias conditions (Fig. 1c). Here we adapted signal detection theory for testing recognition memory in rats to address three questions. First, can recognition memory be assessed in rats using ROC analyses? Second, do rats have distinct processes for recollection and familiarity in recognition memory? Third, is the hippocampus selectively involved in recollection?

We designed a recognition memory task that exploits rats' superb memory capacity for odours. In each daily session, rats initially sampled a list of common household scents. Then, following a 30-min delay, old and new odours were individually presented in a random order. Across sessions the rats' response criterion was biased by altering the difficulty of responding to the test odour and the pay-off ratio for correct 'new' and 'old' responses (see Methods). Initially, we tested 12 intact rats across the full range of

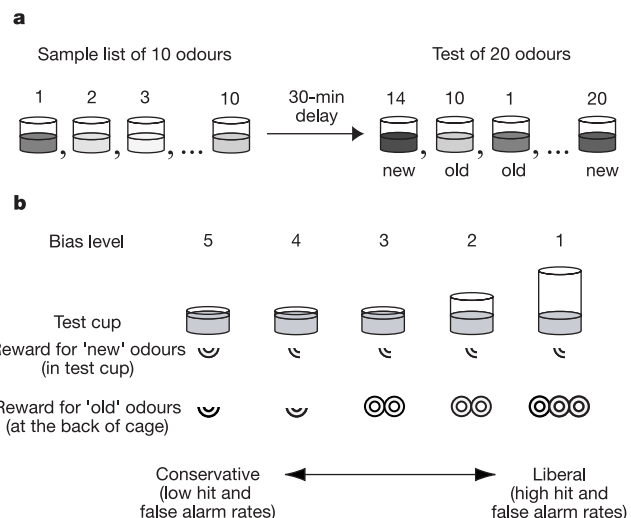
bias levels and then plotted mean scores for each bias level (1–5) (Fig. 1d). The ROC curves were then derived by fitting the data points by using a least-squares model with recollection and familiarity as parameters (see Supplementary Information). The ROC was significantly curvilinear ( $F_{\text{Quad } 2,2} = 30.60$ ,  $P < 0.05$ ), suggesting the contribution of a familiarity process<sup>3,19</sup>. Furthermore, the ROC was asymmetric, with a positive Y-intercept ( $R = 0.40$ ) and the slope of the z-transformed linear approximation was less than 1 ( $t_{11} = -3.17$ ;  $P < 0.05$ ; see Supplementary Information), indicating the presence of a recollection component as well (see ref. 19). This pattern closely matches the overall ROC of humans in verbal recognition performance (compare to Fig. 1a; ref. 20).

On the basis of estimates of recollection and familiarity components of performance calculated for each subject, we then separated the rats into two matched groups (Fig. 1d, inset). One group received selective lesions to the hippocampus and the other received sham control operations (Fig. 2). After recovery, overall control performance was 73% correct and animals with hippocampal lesions were modestly impaired (66%;  $t_{10} = -2.39$ ,  $P < 0.05$ ). The deficit was not attributable to a general shift in response bias (overall response to test cup = 46% for controls, 50% for hippocampal rats;  $t_{10} = 0.766$ ). The ROC of control rats continued to reflect both the recollective and familiarity components (Fig. 1e). By contrast, the ROC of animals with selective hippocampal lesions became fully symmetrical ( $R = 0$ ; z-transformed slope not different from 1,  $t_5 = 0.10$ ) and remained curvilinear ( $F_{\text{Quad } 2,2} = 155.47$ ,  $P < 0.05$ ), characteristic of recognition based solely on familiarity (compare to Fig. 1b). Furthermore, when the contribution of the recollective component was algebraically removed from the ROC of control animals, the resulting curve superimposed onto the ROC of rats with hippocampal lesions (see 'con F' in Fig. 1e; see Supplementary Information).

Additional analyses of the raw scores also provide compelling evidence that recollection is severely impaired in rats with hippocampal lesions, whereas familiarity is intact. The hit rates of the lesioned group were significantly lower than those of the control group ( $F_{1,10} = 6.61$ ,  $P < 0.05$ ; group  $\times$  bias level interaction:  $F_{4,40} = 2.87$ ,  $P < 0.05$ ) but not from those of the con F group ( $F_{1,10} = 0.26$ ; group  $\times$  bias level interaction:  $F_{4,40} = 0.54$ ). False alarm rates did not differ between the groups ( $F_{1,10} = 0.13$ ; group  $\times$  bias level interaction:  $F_{4,40} = 0.14$ ; see Supplementary Information). Subsequent planned comparisons also showed that familiarity estimates did not differ between the two groups ( $t_{10} = 0.65$ ; Fig. 1e inset; see Supplementary Information) and remained significantly greater



**Figure 2** Lesions of the hippocampus reconstructed on coronal sections of the rat brain. Numbers indicate anterior–posterior (AP) distances (in mm) from bregma. Dark grey, smallest lesion; light grey, largest lesion.



**Figure 3** Odour recognition task. **a**, Sequence of odour presentations. **b**, Test cup heights and reward pay-offs for each bias level.

than 0 (controls:  $t_5 = 4.12$ ,  $P < 0.05$ ; hippocampus:  $t_5 = 5.04$ ,  $P < 0.05$ ), whereas recollection estimates for the hippocampal group were significantly lower than those of the control group ( $t_{10} = 2.13$ ,  $P < 0.05$ ) and did not differ significantly from 0 ( $t_5 = 1.70$ ; controls  $> 0$ :  $t_5 = 3.37$ ,  $P < 0.05$ ).

Given that any performance deficit must result in an ROC closer to the diagonal (chance performance), it is possible that the alteration in the ROC pattern of hippocampus-damaged animals is not specific to recollection, but reflects a generalized decline in memory. To test this hypothesis, we challenged control rats by increasing the memory delay to 75 min. This manipulation succeeded in reducing their overall performance to 64% correct, numerically below that of the rats with hippocampal lesions at the 30-min delay. Under those conditions, the ROC of controls became more linear ( $R^2_{\text{linear } 30 \text{ min}} = 0.86$ ,  $R^2_{\text{linear } 75 \text{ min}} = 0.99$ ;  $t_5 = -3.36$ ,  $P < 0.05$ ), suggesting that familiarity contributed substantially less to performance. However, unlike the ROC of animals with hippocampal lesions, the ROC of controls continued to have an asymmetrical recollective component (Y-intercept  $> 0$ ;  $t_5 = 2.13$ ,  $P < 0.05$ ; z-transformed slope  $< 1$ ,  $t_5 = -4.49$ ,  $P < 0.05$ ; Fig. 1f). The degree of curvature ( $R^2_{\text{Quad}} - R^2_{\text{Linear}}$ ) was also higher in the hippocampal group at 30 min than the control group at 75 min ( $t_{10} = 2.71$ ,  $P < 0.05$ ), confirming that the curves are qualitatively different. This pattern of performance indicates that the deficit observed in rats with hippocampal damage is specific to recollection and not the consequence of a general decrease in performance. Furthermore, these findings are consistent with previous reports that familiarity decays more than recollection shortly after learning in humans<sup>21,22</sup>.

An alternative model of recognition memory argues that an asymmetrical ROC curve can be explained by the combination of two parameters: the difference in memory strength between old and new item distributions ( $d'$ ), and a larger variance of the old item distribution than that for the new items ( $V_{\text{old}} > V_{\text{new}} = 1$ ; ref. 18). The unequal-variance model can account for the ROC of normal rats before surgery ( $V_{\text{old}} = 1.27$  and  $d' = 1.76$ ) and the postoperative ROC of control rats ( $V_{\text{old}} = 1.33$  and  $d' = 1.63$ ). This model can also account for the performance of rats with hippocampal lesions, but the variances become equal ( $V_{\text{old}} = 0.97$ ; not significantly different from  $V_{\text{new}} = 1.0$ ,  $t_5 = 0.63$ ) leaving a signal-detection process governed by only a single variable ( $d' = 1.01$ ), as in the case of the dual-process model. Thus, the unequal-variance model confirms that hippocampal lesions selectively eliminate the specific factor that mediates the identification of old items under the most conservative response criterion, analogous to the effect of removing recollection in the dual-process model.

Our results could explain the pattern of findings from previous studies on monkeys and rats performing the delayed non-match to sample (DNMS) recognition task<sup>10–15,23</sup>. The behavioural protocol used here is quite similar to the DNMS task, and indeed revealed the typical modest impairment in recognition following hippocampal lesions. The present analyses allow us to understand the modest overall deficit as reflecting the combination of a severe and selective impairment in recollection contrasted with normal familiarity. Also, the present results suggest that mixed findings across different types of recognition task<sup>10–17</sup> may be a consequence of differential loading of recollective and familiarity memory demands.

Our findings do not shed light on the subjective experience of recollection and familiarity in animals<sup>24</sup>, but do provide the first objective evidence of a distinction between the two processes in a non-human species. Furthermore, our results also reveal that the threshold retrieval process characteristic of human episodic recollection is dependent on the hippocampus. Other studies of recollection in humans and hippocampal-dependent memory in animals have emphasized the importance of memory for associations between items<sup>4,25,26</sup>, for spatial<sup>16,17</sup> and temporal context<sup>27,28</sup>, for bridging temporal gaps<sup>29</sup> and for the flow of events in unique

experiences<sup>30</sup>. The present observations add to that list of properties a fundamental role for the hippocampus in retrieval dynamics associated with recollective experience. □

## Methods

Subjects were 225–250-g male Long Evans rats maintained at a minimum of 85% of normal body weight. They were initially trained to dig for quartered Cheerio cereal rewards in 125-ml plastic cups filled with playground sand scented with distinct odours (see ref. 30 for details), which were presented individually in the front of the home cage. Then the rats were trained on a non-matching task in which a single odour was presented as a sample, followed by two test odours presented consecutively. On each test, the animal obtained an additional reward by digging in the test cup if the odour was 'new' (that is, non-match) or by refraining from digging in the test cup and approaching an alternate empty cup at the back of the cage if the odour was 'old' (that is, match). Each animal was trained to a criterion of 80% correct over three consecutive sessions (six trials per session; average sessions to criterion =  $11.25 \pm 2.83$  s.d.). In subsequent training, each session consisted of a presentation of a list of five sample odours followed by five new and five old test odours presented in a random order. Subjects reached a criterion of 80% correct over three consecutive sessions (average sessions to criterion =  $37.83 \pm 6.32$ ).

For the final training, five different response criteria (conservative to liberal) were generated by biasing the animal's choices through varying the height of the test cup (4 cm, 6.5 cm and 8.5 cm) and the amount of Cheerios provided at the correct choice (between a quarter and three pieces; see Fig. 3). Initially, each session consisted of a 10-odour list at bias level two (same as previous training) for  $21.75 \pm 2.52$  sessions, then animals were acclimated to the five bias levels for 15 sessions (three per bias level). ROC analyses were performed on the subsequent 20 preoperative and 20 postoperative sessions (four sessions per bias level for each) using a 30-min delay between the end of the sample list and the beginning of the tests. In each session, the first four test odours allowed the animal to set its response criterion, then performance on the remaining 16 test odours was used to generate ROC curves for each subject. In addition, estimates of recollection and familiarity were calculated on the basis of the Y-intercepts and the  $d'$ s (transformed into a probability to facilitate comparison with recollection, see Supplementary Information), respectively, for individual subjects. Also, an idealized ROC curve for familiarity alone was calculated by algebraically removing the estimated recollection component from the control curve (compare with Fig. 1c; see Supplementary Information). Finally, controls were again tested for 20 sessions with a 75-min delay.

Following preoperative training, animals were operated on and received either selective lesions to the hippocampus or sham lesions. Anaesthesia was administered using nitrous oxide and was supplemented by 1% halothane. Atropine sulphate (0.081 mg) was injected to prevent respiratory difficulties, and body temperature was maintained at 37 °C. At each of 12 sites bilaterally, the dura was pierced, a 100- $\mu$ m nichrome electrode (0.7-mm uninsulated tip) was lowered into the hippocampus, and lesions were made by passing a 7–11-mA radiofrequency current (Radionics RFG-4A) for 1 min. Sham controls underwent the same surgery, except that the electrode was not lowered into the brain after puncturing the dura. Animals recovered for two weeks and regained their preoperative weights. Lesioned animals lost 32–61% (mean = 42%) of the total volume of the hippocampus, sparing the adjacent subiculum and perirhinal, postrhinal and entorhinal cortices. The cortex overlying the hippocampus was slightly damaged in some animals. Two animals had slight damage to the optic tract and one to the medial geniculate nucleus.

Received 6 May; accepted 12 July 2004; doi:10.1038/nature02853.

- Yonelinas, A. P. The nature of recollection and familiarity: a review of 30 years of research. *J. Mem. Lang.* **46**, 441–517 (2002).
- Sherman, S. J., Atri, A., Hasselmo, M. E., Stern, C. E. & Howard, M. W. Scopolamine impairs human recognition memory: data and modeling. *Behav. Neurosci.* **117**, 526–539 (2003).
- Yonelinas, A. P. et al. Effects of extensive temporal lobe damage or mild hypoxia on recollection and familiarity. *Nature Neurosci.* **5**, 1236–1241 (2002).
- Davachi, L. & Wagner, A. D. Hippocampal contributions to episodic encoding: insights from relational and item-based learning. *J. Neurophysiol.* **88**, 982–990 (2002).
- Ranganath, C., Johnson, M. K. & D'Esposito, M. Prefrontal activity associated with working memory and episodic long-term memory. *Neuropsychologia* **41**, 378–389 (2003).
- Manns, J. R., Hopkins, R. O., Reed, J. M., Kitchener, E. G. & Squire, L. R. Recognition memory and the human hippocampus. *Neuron* **37**, 171–180 (2003).
- Stark, C. E. & Squire, L. R. Functional magnetic resonance imaging (fMRI) activity in the hippocampal region during recognition memory. *J. Neurosci.* **20**, 7776–7781 (2000).
- Eichenbaum, H., Otto, T. & Cohen, N. J. Two functional components of the hippocampal memory system. *Behav. Brain Sci.* **17**, 449–472, 472–518 (1994).
- Brown, M. W. & Aggleton, J. P. Recognition memory: what are the roles of the perirhinal cortex and hippocampus? *Nature Rev. Neurosci.* **2**, 51–61 (2001).
- Beason-Held, L. L., Rosene, D. L., Killiany, R. J. & Moss, M. B. Hippocampal formation lesions produce memory impairment in the rhesus monkey. *Hippocampus* **9**, 562–574 (1999).
- Zola, S. M. et al. Impaired recognition memory in monkeys after damage limited to the hippocampal region. *J. Neurosci.* **20**, 451–463 (2000).
- Nemanic, S., Alvarado, M. C. & Bachevalier, J. The hippocampal/parahippocampal regions and recognition memory: insights from visual paired comparison versus object-delayed nonmatching in monkeys. *J. Neurosci.* **24**, 2013–2026 (2004).
- Mumby, D. G. Perspectives on object recognition memory following hippocampal damage: lessons from studies on rats. *Behav. Brain Res.* **127**, 159–181 (2001).
- Dudchenko, P. A., Wood, E. R. & Eichenbaum, H. Neurotoxic hippocampal lesions have no effect on odor span and little effect on odor recognition memory but produce significant impairments on spatial span, recognition, and alternation. *J. Neurosci.* **20**, 2964–2977 (2000).

15. Steckler, T., Drinkenburg, W. H., Sahgal, A. & Aggleton, J. P. Recognition memory in rats II. Neuroanatomical substrates. *Prog. Neurobiol.* **54**, 313–332 (1998).
16. Mumby, D. G., Gaskin, S., Glenn, M. J., Schramek, T. E. & Lehmann, H. Hippocampal damage and exploratory preferences in rats: memory for objects, places, and contexts. *Learn. Mem.* **9**, 49–57 (2002).
17. Eacott, M. J. & Norman, G. Integrated memory for object, place, and context in rats: a possible model of episodic-like memory? *J. Neurosci.* **24**, 1948–1953 (2004).
18. Macmillan, N. A. & Creelman, C. D. *Detection Theory: A User's Guide* (Cambridge Univ. Press, New York, 1991).
19. Yonelinas, A. P., Kroll, N. E., Dobbins, I., Lazzara, M. & Knight, R. T. Recollection and familiarity deficits in amnesia: convergence of remember-know, process dissociation, and receiver operating characteristic data. *Neuropsychology* **12**, 323–339 (1998).
20. Yonelinas, A. P. Components of episodic memory: the contribution of recollection and familiarity. *Phil. Trans. R. Soc. Lond. B* **356**, 1363–1374 (2001).
21. Hockley, W. E. Item versus associative information: further comparison of forgetting rates. *J. Exp. Psychol. Learn. Mem. Cogn.* **18**, 1321–1330 (1992).
22. Yonelinas, A. P. & Levy, B. J. Dissociating familiarity from recollection in human recognition memory: different rates of forgetting over short retention intervals. *Psychon. Bull. Rev.* **9**, 575–582 (2002).
23. Murray, E. A. & Mishkin, M. Object recognition and location memory in monkeys with excitotoxic lesions of the amygdala and hippocampus. *J. Neurosci.* **18**, 6568–6582 (1998).
24. Tulving, E. Episodic memory: from mind to brain. *Annu. Rev. Psychol.* **53**, 1–25 (2002).
25. Bunsey, M. & Eichenbaum, H. Conservation of hippocampal memory function in rats and humans. *Nature* **379**, 255–257 (1996).
26. Day, M., Langston, R. & Morris, R. G. Glutamate-receptor-mediated encoding and retrieval of paired-associate learning. *Nature* **424**, 205–209 (2003).
27. Wood, E. R., Dudchenko, P. A., Robitsek, R. J. & Eichenbaum, H. Hippocampal neurons encode information about different types of memory episodes occurring in the same location. *Neuron* **27**, 623–633 (2000).
28. Kesner, R. P. in *Neurobiology of Comparative Cognition* (eds Kesner, R. P. & Olton, D. S.) 179–203 (Lawrence Erlbaum, New Jersey, 1990).
29. Clark, R. E. & Squire, L. R. Classical conditioning and brain systems: the role of awareness. *Science* **280**, 77–81 (1998).
30. Fortin, N. J., Agster, K. L. & Eichenbaum, H. Critical role of the hippocampus in memory for sequences of events. *Nature Neurosci.* **5**, 458–462 (2002).

Supplementary Information accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank J. O'Connell, A. Milewski, L. Giocomo, J. Estes, D. Tosa, R. Kline, J. Davidson and B. Goldberg for help with behavioural testing. We also thank R. Komorowski for histological work, A. Yonelinas for providing the least-squares curve-fitting algorithm and J. Manns for comments on an earlier version of this manuscript. Supported by NIA and NIMH.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to H.B.E. (hbe@bu.edu).

## Restricted growth of Schwann cells lacking Cajal bands slows conduction in myelinated nerves

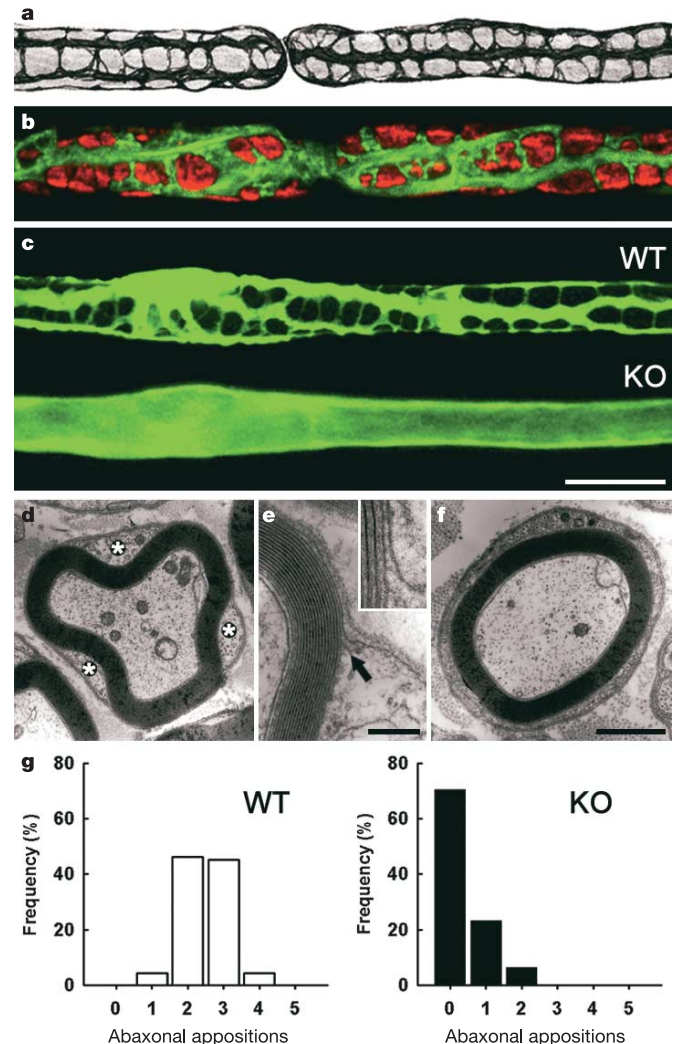
Felipe A. Court, Diane L. Sherman, Thomas Pratt, Emer M. Garry, Richard R. Ribchester, David F. Cottrell, Susan M. Fleetwood-Walker & Peter J. Brophy

Centre for Neuroscience Research, University of Edinburgh, Edinburgh EH9 1QH, UK

Nerve impulses are propagated at nodes of Ranvier in the myelinated nerves of vertebrates. Internodal distances have been proposed to affect the velocity of nerve impulse conduction<sup>1</sup>; however, direct evidence is lacking, and the cellular mechanisms that might regulate the length of the myelinated segments are unknown. Ramón y Cajal described longitudinal and transverse bands of cytoplasm or trabeculae in internodal Schwann cells and suggested that they had a nutritive function<sup>2</sup>. Here we show that internodal growth in wild-type nerves is precisely matched to nerve extension, but disruption of the cytoplasmic bands in Periaxin-null mice impairs Schwann cell elongation

during nerve growth. By contrast, myelination proceeds normally. The capacity of wild-type and mutant Schwann cells to elongate is cell-autonomous, indicating that passive stretching can account for the lengthening of the internode during limb growth. As predicted on theoretical grounds, decreased internodal distances strikingly decrease conduction velocities and so affect motor function. We propose that microtubule-based transport in the longitudinal bands of Cajal permits internodal Schwann cells to lengthen in response to axonal growth, thus ensuring rapid nerve impulse transmission.

Nodes of Ranvier in peripheral nerves are flanked by Schwann



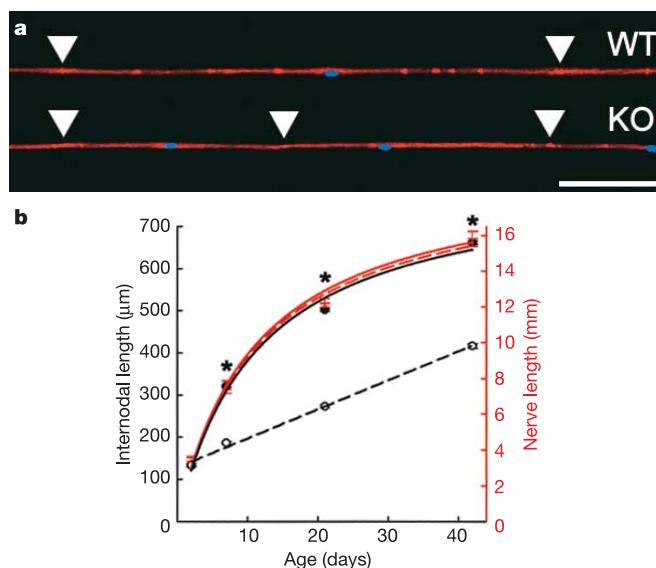
**Figure 1** Longitudinal and transverse bands of Cajal in Schwann cells and their disruption in quadriceps nerve of Periaxin-null (KO) mice at 3 weeks. **a**, Longitudinal and transverse protoplasmic bands stained with silver by Ramón y Cajal (reproduced with permission)<sup>2</sup>. **b**, Teased fibres double-labelled with TRITC-phalloidin (green) and an antibody against DRP2 (red). Schwann cell cytoplasm is excluded from spheroidal clusters immunopositive for DRP2. **c**, Immunostaining of fibres from WT and *Prx* KO mice for the Schwann cell cytoplasmic protein S100. Scale bar, 20  $\mu$ m in **b** and **c**. **d–f**, Electron micrographs of transverse sections of WT and KO quadriceps nerves. **d**, WT Schwann cell cytoplasm (asterisks) is restricted to regions delimited by appositions between the Schwann cell plasma membrane and the abaxonal layer of the myelin sheath. **e**, The sharp transition between the apposition and cytoplasmic zones is shown by the arrow. Scale bar, 0.2  $\mu$ m. The inset shows a high-power view of the transition zone. **f**, In the absence of appositions in the KO, the Schwann cell cytoplasm forms a concentric ring around the myelin sheath. Scale bar, 1  $\mu$ m (**d** and **f**). **g**, The proportion of abaxonal appositions per Schwann cell in WT and KO mice ( $n = 3$  for WT and KO).



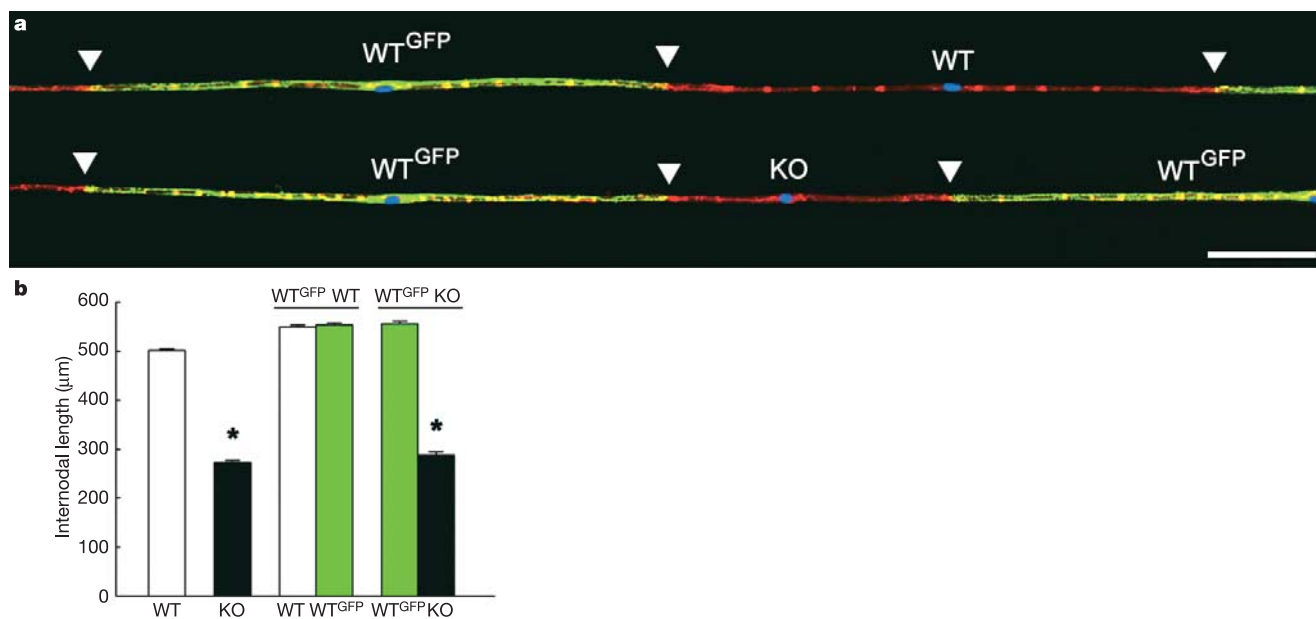
cells, and in theory the length of the internodal Schwann cell should influence the rate of impulse conduction<sup>1</sup>. However, this has not been experimentally verified, largely because the two other main influences on conduction velocity, namely axon diameter and myelin thickness, normally co-vary with internodal length. Further, it is not known what mechanisms set the internodal distance and allow Schwann cells to elongate so extensively during nerve growth *in vivo*. The cytoplasmic meshwork beneath the plasma membrane of myelinating Schwann cells was first described by Ramón y Cajal (Fig. 1a), who speculated that these longitudinal and transverse protoplasmic trabeculae might have a nutritive function; however, their role in cell growth and maturation has remained unclear<sup>2</sup>.

We have previously identified spheroidal domains in the Schwann cell plasma membrane that contain the L-Periaxin-Dystrophin-related protein 2 (DRP2)–Dystroglycan (PDG) complex<sup>3</sup>. Double-labelling for DRP2 and cytoplasmic microfilaments by immunofluorescence showed that the PDG-rich domains are surrounded by cytoplasmic bands (Fig. 1b). To address the function of the bands, we first determined whether they were deranged in Periaxin-null (*Prx*<sup>−/−</sup>) mice, because we had previously shown that the PDG complex is biochemically disrupted in these mice<sup>3</sup>. Immunostaining for the Schwann cell cytoplasmic marker S100 revealed that the cytoplasmic bands were absent from the Schwann cells of *Prx*<sup>−/−</sup> mice (Fig. 1c). Electron microscopy showed that there were no appositions between the abaxonal surface of the myelin sheath and the plasma membrane of the *Prx*<sup>−/−</sup> Schwann cells (Fig. 1d–g), and this was consistent with our previous ultrastructural observation that DRP2 and Periaxin localize together at these appositions<sup>3</sup>. Instead of being compartmentalized in bands, the Schwann cell cytoplasm now formed a continuous annulus under the plasma membrane (Fig. 1f).

What are the consequences of disrupting the longitudinal Cajal bands, in the light of Ramón y Cajal's proposal that they might serve a nutritive function? Although inactivation of the *Prx* gene causes

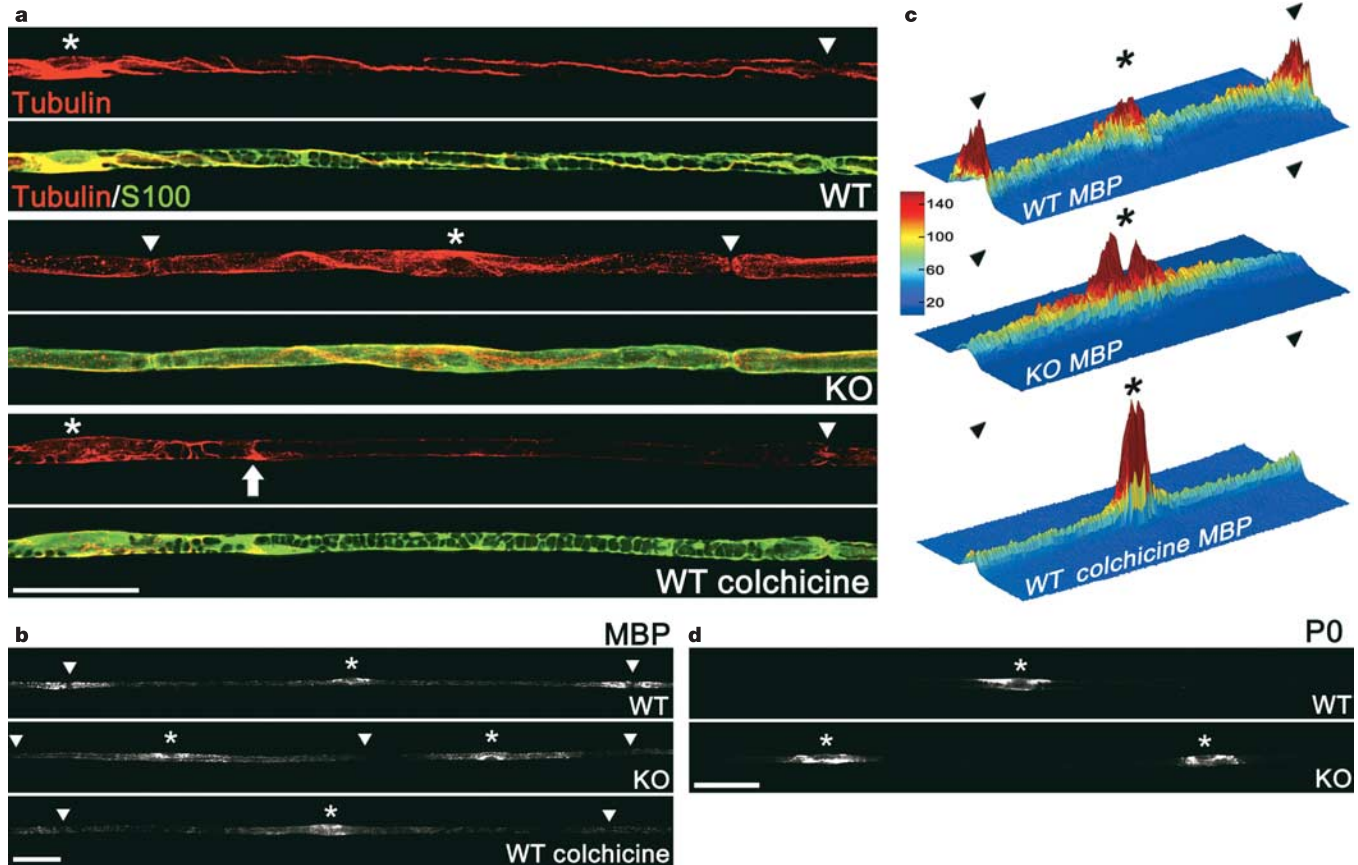


**Figure 2** Internodal length is decreased in Schwann cells from *Prx*<sup>−/−</sup> mice lacking Cajal bands. **a**, Teased fibres from 3-week-old WT and KO quadriceps nerves were stained with TRITC-phalloidin (red) and DAPI (blue). Nodes of Ranvier are indicated by arrowheads. Scale bar, 100 μm. **b**, The growth rate of WT and KO internodal lengths fitted a rectangular hyperbola and a straight line respectively. The growth rates of WT and KO nerve length were similar and both fitted rectangular hyperbolas (the null hypothesis that there was no significant difference between these fitted curves was supported by *F*-test; *P* = 0.945). Internodal lengths of WT (continuous black line) and KO (dashed black line) are comparable at 2 days. Thereafter, internodal growth of WT Schwann cells matches the increase in length of the quadriceps nerve (continuous red line) exactly. In contrast, the rate of internodal growth in the KO is decreased, even though the nerve grows at a normal rate (dashed red line). Values are means ± s.e.m. for three animals (asterisks, *P* < 0.0001; Student's *t*-test for internodal lengths).



**Figure 3** The capacity of Schwann cells to specify internodal length is cell-autonomous. Teased fibres in quadriceps nerves from 3-week-old mice were stained with TRITC-phalloidin and DAPI. **a**, Nerves were from chimaeric mice containing GFP-tagged Schwann cells (S129 background) and either WT Schwann cells or *Prx*KO Schwann cells (both C57BL/6 background). The green signal is from GFP. Nodes of Ranvier are indicated by arrowheads. Scale bar, 100 μm. **b**, The first two bars show that internodal lengths from non-chimaeric KO mice are decreased (\**P* < 0.0001 by Student's *t*-test; *n* = 3). In WT<sup>GFP</sup>–WT chimaeras, both types of Schwann cells had similar internodal lengths. In

WT<sup>GFP</sup>–KO chimaeras, WT<sup>GFP</sup> Schwann cells had similar internodal lengths to those in WT<sup>GFP</sup>–WT chimaeras, whereas KO Schwann cells had decreased internodal lengths (asterisks, *P* < 0.0001 by Student's *t*-test; *n* = 3). The internodes of Schwann cells in WT<sup>GFP</sup>–WT chimaeras were significantly larger than those of C57BL/6 WT (*P* < 0.0001 by Student's *t*-test), but the internodal lengths of KO Schwann cells were decreased similarly in a chimaeric environment and in a C57BL/6 environment (*P* = 0.07 by Student's *t*-test). All values are means + s.e.m.



**Figure 4** Intact Cajal bands and microtubules are required for MBP mRNA localization at the paranodes. Tubulin (red) and the Cajal band marker S100 (green) were visualized by immunofluorescence, and MBP and P0 mRNA were detected by *in situ* hybridization (images photographically inverted for clarity) in teased sciatic nerve fibres from 3-week-old WT and KO nerves with digoxigenin-labelled probes. **a**, In WT Schwann cells tubulin localizes with S100 in the Cajal bands, whereas in the absence of the bands tubulin staining becomes punctate as the microtubules approach the paranodes. Complete depolymerization of WT Schwann cell microtubules with colchicine *in vivo* followed by recovery for 3 days allows the microtubules to reform and partly extend (arrow indicates

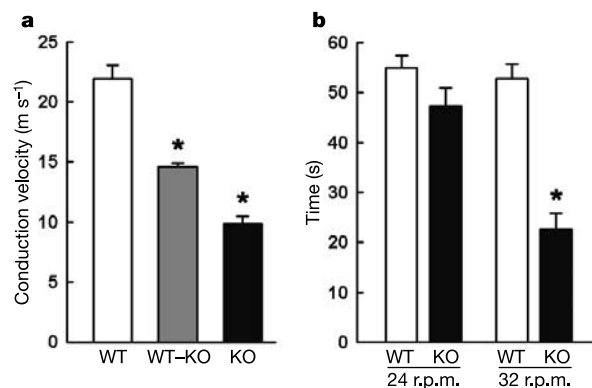
extent of growth). Note that Cajal bands are unaffected. **b**, MBP mRNA in WT Schwann cells accumulates in the perinuclear and paranodal domains, whereas in KO Schwann cells lacking Cajal bands the strong MBP mRNA signal declines from the perinuclear region towards the paranodes. Depolymerization of microtubules in WT cells prevents the concentration of MBP mRNA at the paranodes. Scale bar, 50  $\mu$ m. **c**, Pseudocolour surface plot of the mean internodal signal intensity (scale, 0–255) for MBP mRNA in WT ( $n = 9$ ), KO ( $n = 12$ ) and colchicine-treated WT Schwann cells ( $n = 10$ ). **d**, P0 mRNA is localized to the perinuclear region in both WT and KO Schwann cells. Nuclei and nodes of Ranvier are indicated by asterisks and arrowheads respectively. Scale bar, 50  $\mu$ m.

demyelination in mature mice<sup>4</sup>, the myelin sheath of murine quadriceps nerve fibres was of normal thickness at 3 weeks as measured by the *g*-ratio (axon diameter/fibre diameter) (wild type (WT),  $0.62 \pm 0.01$ ; *Prx*<sup>-/-</sup>,  $0.63 \pm 0.01$ ), and the mean fibre diameter was unchanged (WT,  $6.73 \pm 0.06 \mu$ m; *Prx*<sup>-/-</sup>,  $6.67 \pm 0.07 \mu$ m). Furthermore, there seemed to be no derangements to the protein content, protein localization or ultrastructural organization of the nodal or paranodal apparatus at 3 weeks (Supplementary Fig. 1).

In spite of the apparently normal assembly of myelin by Schwann cells, but consistent with Ramón y Cajal's proposal, there was a striking decrease in their longitudinal growth (Fig. 2a). At post-natal-day 2 (P2), when myelinating Schwann cells have established a 1:1 relationship with segments of quadriceps nerve axons, WT and knockout (KO) Schwann cells were the same length (Fig. 2b). WT Schwann cells then elongated at the same rate as the growing nerve, whereas Periaxin-null Schwann cells had lower growth rates; this deficit persisted into adulthood in spite of the fact that WT and KO quadriceps nerves grew to the same length (Fig. 2b). Consistent with the absence of spontaneous firing in their quadriceps nerves, we never observed naked axonal segments in KO mice. Hence, during the early rapid growth phase of the nerve, extra Schwann cells must be needed to ensheath nerves in the mutant, although their origin remains to be determined. At later phases after

P21 it seems that KO Schwann cells can match the slower rates of nerve growth.

These results prompted us to use chimaeric mice to test whether the ability of Schwann cells to elongate in response to axonal growth was cell-autonomous. Strain S129-derived embryonic stem (ES) cells tagged with tau-green fluorescent protein (GFP)<sup>5</sup> were injected into blastocysts to track the fate of ES cells carrying WT *Prx* alleles. We had previously shown that the GFP fusion transgene is robustly expressed in myelinating Schwann cells. To confirm that these ES cells developed into Schwann cells with normal internodal lengths, we injected them into C57BL/6 WT as well as KO blastocysts. In each case, peripheral nerves contained Schwann cells derived from both endogenous and injected stem cells (Fig. 3a). In contrast to the normal elongation of Schwann cells tagged with GFP in a WT or KO environment, KO Schwann cells were unable to elongate normally whether alone or flanked by GFP-tagged Schwann cells with normal longitudinal bands and WT internodal lengths (Fig. 3b). This indicated that the capacity to elongate was a cell-autonomous property of Schwann cells. Internodal lengths in chimaeric nerves containing both GFP-tagged Schwann cells (derived from strain S129 mice) and WT Schwann cells (from C57BL/6 mice) were greater than in WT C57BL/6 mice because these chimaeric mice were larger and had longer quadriceps nerves (compare Figs 2b and 3b). In contrast, the internodal lengths of Periaxin-null Schwann



**Figure 5** Peripheral nerve function is compromised in *Prx*<sup>-/-</sup> animals. **a**, Nerve conduction velocities in 3-week-old quadriceps nerves were decreased in KO compared with WT mice, and chimaeras had intermediate values ( $57.1 \pm 3.1\%$  KO Schwann cells,  $n = 3$ ; mean  $\pm$  s.e.m.). Error bars indicate standard error of the regression coefficient (asterisks,  $P < 0.0001$  for comparisons with WT, Student's *t*-test;  $n = 4$  for WT and KO,  $n = 3$  for chimaeras). **b**, A significant decrease in KO motor coordination compared with

cells in a chimaeric or pure C57BL/6 environment were indistinguishable (Fig. 3b). This further indicated that there was an upper limit to the ability of the mutant Schwann cells to elongate, irrespective of the extent of axon growth.

Microtubule-based messenger RNA translocation is believed to permit the incorporation of newly synthesized myelin basic protein (MBP) at the growing extremities of myelin-forming glia<sup>6-9</sup>. To test whether impaired transport of mRNA and proteins from the perinuclear to the peripheral regions of myelin-forming glia as a result of microtubule disruption might contribute to the decreased rate of elongation of Schwann cells lacking Cajal bands, we analysed the distribution of microtubules and MBP mRNA in single teased fibres from WT and KO mice. WT Schwann cells had an organized microtubule network that extended from the nucleus to the paranodes along the Cajal bands, whereas the tubulin staining in cells lacking Cajal bands frequently became punctate as the microtubules approached the paranodes (Fig. 4a). Importantly, the mutant Schwann cells could no longer accumulate MBP mRNA in the paranodal regions (Fig. 4b, c). Furthermore, the distal accumulation of MBP mRNA in WT Schwann cells was prevented when microtubules were reversibly disrupted, as found previously for oligodendrocytes<sup>7</sup> (Fig. 4a-c). Hence, deficits in microtubule-based transport might underlie the decreased capacity of Periaxin-null Schwann cells to elongate. In contrast to MBP mRNA, no discernible differences between WT and KO mice were observed in the distribution of the mRNA encoding the integral membrane protein P0 (Fig. 4d).

The decreased internodal length in KO mice, taken together with their normal axon diameter and myelin thickness, provided a unique opportunity to determine whether nerve conduction velocities are modulated by the internodal distance, as predicted from theory<sup>1</sup>. We measured the conduction time of compound action potentials in quadriceps nerves from 3-week-old WT, mutant and chimaeric mice. Nerve conduction velocities were decreased by more than 50% in KO compared with WT animals, and nerves that contained a mixture of WT and mutant Schwann cells had intermediate values, which was consistent with the effect of internodal length on conduction velocity (Fig. 5a). The substantial decrease in motor nerve conduction velocity in nerves lacking Cajal bands was accompanied by poorer performance on the RotaRod, a robust behavioural test for motor dysfunction (Fig. 5b). In sensory reflex behavioural tests, 3-week-old mice did not display the neuropathic pain behaviour that results from demyelination in mature *Prx*<sup>-/-</sup> mice<sup>4</sup>. They had normal withdrawal latencies in tests for both

**c**

| Response                                               | WT               | KO               |
|--------------------------------------------------------|------------------|------------------|
| Withdrawal latency from noxious heat (s)               | $8.5 \pm 0.6$    | $8.4 \pm 0.5$    |
| Mechanical withdrawal threshold (mN mm <sup>-2</sup> ) | $281.7 \pm 15.5$ | $290.8 \pm 14.3$ |

WT at 3 weeks was observed on the RotaRod at 32 r.p.m. but not at 24 r.p.m. Values are means  $\pm$  s.e.m. (asterisk,  $P < 0.0001$ , Student's *t*-test;  $n = 6$  for WT and KO).

**c**, 3-week-old WT and KO mice showed no significant differences in a cutaneous reflex test of hindpaw withdrawal threshold to mechanical stimulation or in their withdrawal latency to thermal nociceptive sensitivity. Values are means  $\pm$  s.e.m. ( $n = 6$  for WT and KO).

thermal hyperalgesia and mechanical allodynia, confirming the functional intactness of the myelin sheath (Fig. 5c). The relationship between the absence of longitudinal bands, decreased internodal length and impaired conduction in Periaxin-null mice was underlined by an analysis of *CMTX* mice, which also display late-onset demyelination. The quadriceps nerves from 3-week-old *CMTX* mice have normal bands and their Schwann cell lengths and conduction velocities are not decreased in comparison with WT animals (ref. 10 and Supplementary Fig. 2).

The physiological relationships between fibre diameter, myelin thickness and conduction velocity are well established<sup>11</sup>. However, our understanding of how internodal length influences conduction velocity has been largely theoretical<sup>1,12</sup>. We have used a recent model of mammalian motor nerve fibres that includes detailed morphological and electrical parameters<sup>13</sup>. As found in earlier theoretical treatments<sup>1</sup>, we found a substantial decrease in the rate of conduction when changes in the internodal distance were modelled within our experimental range from 500  $\mu$ m down to 250  $\mu$ m (Supplementary Fig. 3). Nerve conduction velocities therefore become much more sensitive to changes in internodal lengths in the shorter ranges.

Ranvier speculated that internodal length and nerve elongation are related during body growth because internodes are longer in larger animals<sup>14</sup>. From our studies, the ability of myelinating Schwann cells to elongate during the postnatal development of peripheral nerves was cell-autonomous but matched the growth rate of axons precisely, thus supporting the proposal that passive stretching could account for the establishment of internodal lengths during nerve growth<sup>15,16</sup>. Schwann cells lacking longitudinal bands were unable to keep pace with axon growth, and the consequence of this decreased capacity for elongation was that the internodal distances were shorter and conduction velocities were slowed in mutant nerves. We conclude that Schwann cell elongation is an essential feature of the functional development of the vertebrate nervous system. Finally, we propose that the longitudinal cytoplasmic structures in Schwann cells be named Cajal bands to acknowledge both Ramón y Cajal's discovery and our growing understanding of their function. □

## Methods

### Materials and antibodies

Primary antibodies and dyes for microscopy were used at the following dilutions and concentrations; rabbit anti-DRP2 (ref. 3), 1:200; rabbit anti-S100 (Sigma), 1:200; rabbit anti-neurofascin<sup>17</sup>, 1:1000; guinea pig anti-Caspr<sup>17</sup>, 1:200; mouse anti-dystroglycan<sup>3</sup>, 1:100; mouse anti-pan-Na<sup>+</sup> channel (gift from M. Rasband), 1:300; rat anti-tubulin



(YL1/2; Serotec), 1:200; chicken anti-betaIV-spectrin<sup>18</sup> (gift from M. Komada), 1:100; 4,6-diamidino-2-phenylindole (DAPI; Sigma), 4 µg ml<sup>-1</sup>; tetramethylrhodamine β-isothiocyanate (TRITC)-phalloidin (Sigma), 50 ng ml<sup>-1</sup>. The secondary antibodies were fluorescein isothiocyanate (FITC)-conjugated anti-rabbit IgG, 1:200; TRITC-donkey anti-guinea-pig IgG, 1:200; FITC-donkey anti-chicken IgY, 1:100 (all from Jackson ImmunoResearch); FITC-goat anti-mouse IgG1, 1:200 (Southern Biotech); Alexa Fluor-goat anti-rabbit IgG, 1:100 (Molecular Probes). For western blotting<sup>19</sup>, primary antibodies were used at tenfold greater dilutions.

### Microscopy and morphometry

Depolymerization of Schwann cell microtubules in sciatic nerve *in vivo* was performed by topical treatment with colchicine as described<sup>20</sup>, and microtubules were allowed to reform for 3 days. By 5 days after colchicine treatment the microtubules had reached the paranodes showing that microtubule disruption is completely reversible (data not shown). Teased fibres were prepared from nerves fixed for 1 h in 4% paraformaldehyde, 0.1 M sodium phosphate buffer pH 7.3, and washed in several changes of phosphate buffer. The lengths of quadriceps nerves were measured from spinal cord exit to muscle insertion point. Teased fibres from quadriceps nerves of WT and KO mice were stained with TRITC-phalloidin and DAPI, and 100 internodes and the lengths of two quadriceps nerves were measured for each animal ( $n = 3$ ). For chimaeras, 200 internodal distances were measured for each Schwann cell type ( $n = 3$ ). Nerves were prepared for electron microscopy as described previously<sup>4</sup>. For measurement of g-ratios, micrographs of randomly selected fields of ultrathin transverse sections of quadriceps nerves from 3-week-old WT and KO mice ( $n = 3$  for each group, 120 fibres per group) were scanned and analysed using NIH Image. The g-ratio was calculated and all results are shown as means  $\pm$  s.e.m. The best fit for the growth rates of internodal and nerve length was found ( $F$ -test; Prism 3.03) and superimposed on the measured values shown in Fig. 2b. The growth rate of WT internodal length fitted a rectangular hyperbola, whereas the growth rate of KO internodal length fitted a straight line (linear regression). The growth rates of WT and KO nerve length both fitted a rectangular hyperbola. Bright-field images of teased sciatic nerve fibres stained for P0 and MBP mRNA by *in situ* hybridization with digoxigenin-labelled probes<sup>21</sup> were photographically inverted to compare the signal intensity between WT, KO and colchicine-treated WT Schwann cells. Sense probes gave negligible background staining. Images of internodes were normalized by rescaling to a length of 400 pixels and imported into MATLAB software as a two-dimensional matrix representing pixel intensity. A three-dimensional array was created by concatenating individual images in the Z dimension. From this array, the mean value of pixel intensity in the Z dimension was calculated for each pixel, resulting in a two-dimensional matrix.

### Electrophysiology

Quadriceps nerves from 3-week-old KO and WT mice were transferred from oxygenated Krebs solution to an isolated chamber containing an array of Ag/AgCl electrodes with 1-mm intervals and surrounded by liquid paraffin maintained at 37 °C for periods no longer than 10 min. The proximal end of the nerve was excited by a square wave (0.1 ms, 0.1–1.5 V) and the conduction distance was varied from 2 mm to 7 mm by altering the stimulating electrode position. The voltage was adjusted to ensure exact duplication of the active population, and the compound action potential was viewed on a storage oscilloscope. Values were stored as digitized signals with the use of Chart software (MacLab System). Conduction times were measured as described<sup>4</sup>.

### Behavioural testing

Mice were conditioned to the RotaRod 1 day before the trial, and the RotaRod test was terminated either when the mouse fell from the rod or at 60 s. Four trials per mouse were performed, separated by at least 10 min to avoid exhaustion of the animal. The thresholds and times for hindpaw withdrawal in response to graded mechanical stimulation and thermal stimulus were performed as described<sup>4</sup>.

### Chimaeras

Chimaeras were obtained by injecting the ES cell line E14Tg2aSc4TP6.3 expressing GFP-tagged tau protein into WT or KO blastocysts<sup>5</sup>. For internodal length measurements in quadriceps nerves, chimaeras were selected with an approximately equal contribution from each Schwann cell type.

Received 9 May; accepted 13 July 2004; doi:10.1038/nature02841.

- Brill, M. H., Waxman, S. G., Moore, J. W. & Joyner, R. W. Conduction velocity and spike configuration in myelinated fibres: computed dependence on internode distance. *J. Neurol. Neurosurg. Psychiatry* **40**, 769–774 (1977).
- Ramón y Cajal, S. *Histology* (Bailliere, Tindall & Cox, London, 1933).
- Sherman, D. L., Fabrizi, C., Gillespie, C. S. & Brophy, P. J. Specific disruption of a Schwann cell dystrophin-related protein complex in a demyelinating neuropathy. *Neuron* **30**, 677–687 (2001).
- Gillespie, C. S. *et al.* Peripheral demyelination and neuropathic pain behavior in periaxin-deficient mice. *Neuron* **26**, 523–531 (2000).
- Pratt, T., Sharp, L., Nichols, J., Price, D. J. & Mason, J. O. Embryonic stem cells and transgenic mice ubiquitously expressing a tau-tagged green fluorescent protein. *Dev. Biol.* **228**, 19–28 (2000).
- Ainger, K. *et al.* Transport and localization elements in myelin basic protein mRNA. *J. Cell Biol.* **138**, 1077–1087 (1997).
- Carson, J. H., Worboys, K., Ainger, K. & Barbarese, E. Translocation of myelin basic protein mRNA in oligodendrocytes requires microtubules and kinesin. *Cell Motil. Cytoskeleton* **38**, 318–328 (1997).
- Colman, D. R., Kreibich, G., Frey, A. B. & Sabatini, D. D. Synthesis and incorporation of myelin polypeptides into CNS myelin. *J. Cell Biol.* **95**, 598–608 (1982).
- Griffiths, I. R. *et al.* Expression of myelin protein genes in Schwann cells. *J. Neurocytol.* **18**, 345–352 (1989).

- Anzini, P. *et al.* Structural abnormalities and deficient maintenance of peripheral nerve myelin in mice lacking the gap junction protein connexin 32. *J. Neurosci.* **17**, 4545–4551 (1997).
- Hursh, J. B. Conduction velocity and diameter of nerve fibers. *Am. J. Physiol.* **127**, 131–139 (1939).
- Huxley, A. F. & Stampfli, R. Evidence for saltatory conduction in peripheral myelinated nerve fibres. *J. Physiol. (Lond.)* **108**, 315–339 (1949).
- McIntyre, C. C., Richardson, A. G. & Grill, W. M. Modeling the excitability of mammalian nerve fibers: influence of afterpotentials on the recovery cycle. *J. Neurophysiol.* **87**, 995–1006 (2002).
- Ranvier, L. Des étranglements annulaires et des segments interannulaires chez les Raies et les Torpilles. *C. R. Acad. Sci.* **75**, 1129–1132 (1872).
- Hiscoe, H. B. Distribution of nodes and incisures in normal and regenerated nerve. *Anat. Rec.* **99**, 447–475 (1947).
- Schlaepfer, W. W. & Myers, E. K. Relationship of myelin internode elongation and growth in the rat sural nerve. *J. Comp. Neurol.* **147**, 255–266 (1973).
- Tait, S. *et al.* An oligodendrocyte cell adhesion molecule at the site of assembly of the paranodal axoglial junction. *J. Cell Biol.* **150**, 657–666 (2000).
- Komada, M. & Soriano, P. βIV-spectrin regulates sodium channel clustering through ankyrin-G at axon initial segments and nodes of Ranvier. *J. Cell Biol.* **156**, 337–348 (2002).
- Gillespie, C. S., Sherman, D. L., Blair, G. E. & Brophy, P. J. Periaxin, a novel protein of myelinating Schwann cells with a possible role in axonal ensheathment. *Neuron* **12**, 497–508 (1994).
- Trapp, B. D. *et al.* Polarization of myelinating Schwann cell surface membranes: role of microtubules and the trans-Golgi network. *J. Neurosci.* **15**, 1797–1807 (1995).
- Collinson, J. M., Marshall, D., Gillespie, C. S. & Brophy, P. J. Transient expression of neurofascin by oligodendrocytes at the onset of myelination: implications for mechanisms of axon-glial interaction. *Glia* **23**, 11–23 (1998).

**Supplementary Information** accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank H. Anderson and L. Ferguson for assistance, B. Smith for technical support, S. Scherer for comments, and K. Willeke and T. Ott of Bonn University for the CMTX mice. Figure 1a is reproduced with the permission of the Cajal Institute, CSIC, Madrid, Spain, copyright inheritors of Santiago Ramón y Cajal. This work was supported by the Wellcome Trust.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to P.J.B. ([peter.brophy@ed.ac.uk](mailto:peter.brophy@ed.ac.uk)).

## Calcium transients in astrocyte endfeet cause cerebrovascular constrictions

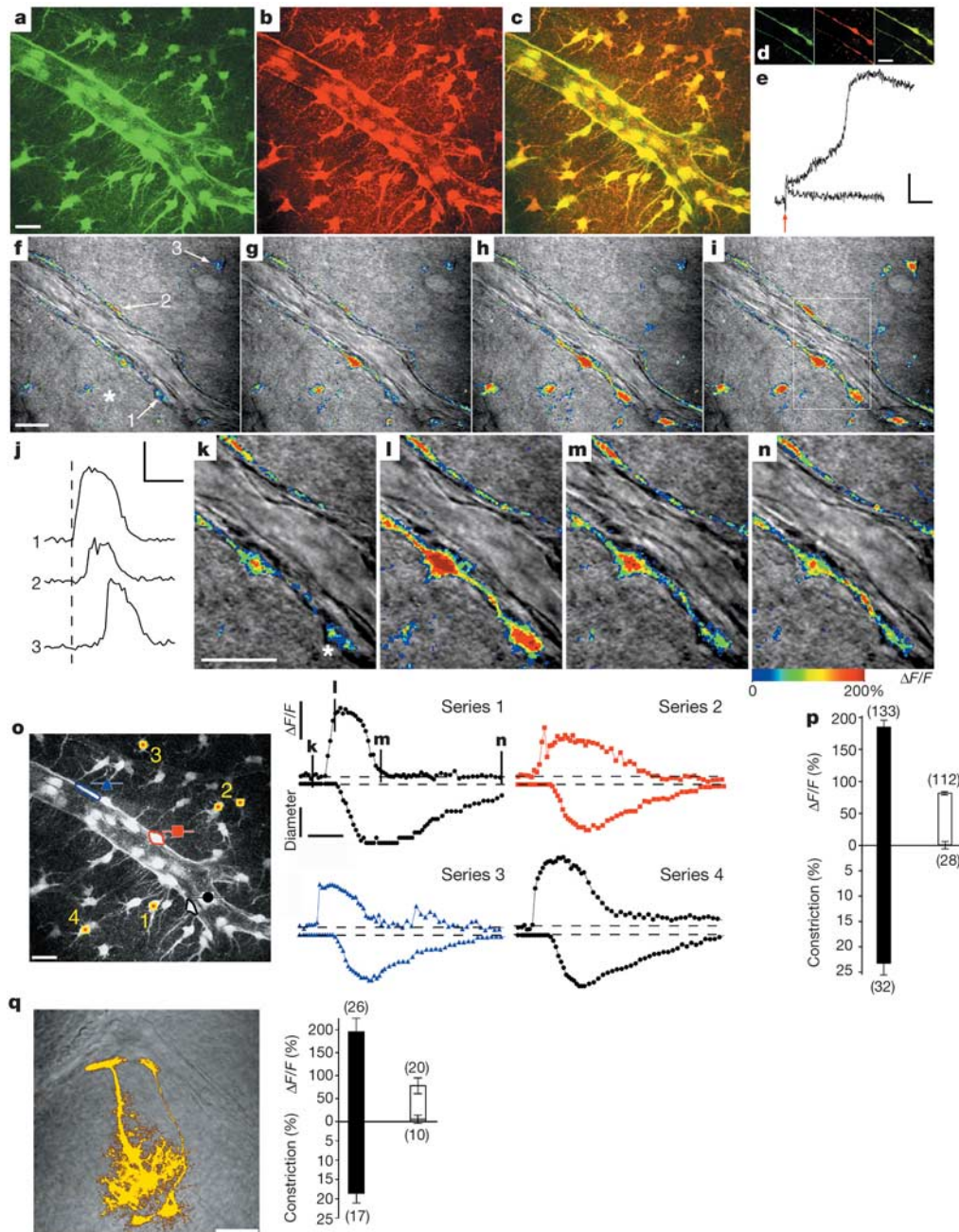
Sean J. Mulligan & Brian A. MacVicar

Brain Research Centre, Department of Psychiatry, University of British Columbia, 2211 Wesbrook Mall, Vancouver, BC, Canada, V6T 2B5

Cerebral blood flow (CBF) is coupled to neuronal activity and is imaged *in vivo* to map brain activation<sup>1</sup>. CBF is also modified by afferent projection fibres that release vasoactive neurotransmitters<sup>2,3</sup> in the perivascular region, principally on the astrocyte endfeet<sup>4,5</sup> that outline cerebral blood vessels<sup>6</sup>. However, the role of astrocytes in the regulation of cerebrovascular tone remains uncertain. Here we determine the impact of intracellular Ca<sup>2+</sup> concentrations ([Ca<sup>2+</sup>]<sub>i</sub>) in astrocytes on the diameter of small arterioles by using two-photon Ca<sup>2+</sup> uncaging<sup>7,8</sup> to increase [Ca<sup>2+</sup>]<sub>i</sub>. Vascular constrictions occurred when Ca<sup>2+</sup> waves evoked by uncaging propagated into the astrocyte endfeet and caused large increases in [Ca<sup>2+</sup>]<sub>i</sub>. The vasoactive neurotransmitter noradrenaline<sup>2,3</sup> increased [Ca<sup>2+</sup>]<sub>i</sub> in the astrocyte endfeet, the peak of which preceded the onset of arteriole constriction. Depressing increases in astrocyte [Ca<sup>2+</sup>]<sub>i</sub> with BAPTA inhibited the vascular constrictions in noradrenaline. We find that constrictions induced in the cerebrovasculature by increased [Ca<sup>2+</sup>]<sub>i</sub> in astrocyte endfeet are generated through the phospholipase A<sub>2</sub>-arachidonic acid pathway and 20-hydroxyeicosatetraenoic acid production. Vasoconstriction by astrocytes is a previously unknown mechanism for the regulation of CBF.

The dynamic modulation of the cerebral microvasculature by astrocytes has been difficult to assess owing to the problem of selectively stimulating astrocytes. Most stimuli that increase  $[Ca^{2+}]_i$  or induce  $Ca^{2+}$  waves in astrocytes (for example, transmitter application or electrical stimulation) could also act directly on the

vascular cells (that is, endothelial cells or smooth muscle cells) or neurons. We have examined the impact of increased  $[Ca^{2+}]_i$  in astrocyte endfeet on the vasculature by selectively increasing  $[Ca^{2+}]_i$  in astrocytes by using highly localized flash photolysis of caged  $Ca^{2+}$ , which is possible only with two-photon laser scanning



**Figure 1** Two-photon photolysis of caged  $Ca^{2+}$  in identified astrocytes initiates  $Ca^{2+}$  waves that propagate to astrocyte endfeet and induce arteriole constriction. **a–d**, Two-dimensional projections showing that GFP-positive astrocytes (**a**), and the endfeet that outline an arteriole were selectively loaded with Rhod-2 (**b**); merged in **c**, single frames in **d**. **e**,  $Ca^{2+}$  uncaging in astrocytes (at red arrow) (in an area of 1 by 5  $\mu m$ , bottom trace; 1 by 10  $\mu m$  area, upper trace). **f–j**, Infrared-transmitted and Rhod-2 fluorescence (pseudocolour) overlay images showing the flash-induced  $Ca^{2+}$  wave (transients shown in **j**) propagates into the astrocyte endfeet at the vessel (labelled 1 and 2 in control, **f**) and precedes vessel constriction (**g–i**). The  $Ca^{2+}$  wave was induced by flashing an astrocyte cell body 10  $\mu m$  below the image plane (asterisk). **k–n**, Images from the area outlined in **i** showing resting  $Ca^{2+}$  and vessel diameter under resting conditions (**k**), high endfoot  $[Ca^{2+}]_i$  preceding constriction (**l**), maximal constriction/recovered  $Ca^{2+}$  (**m**) and recovery

(**n**). **o**, Repeated arteriole constrictions (series 1 to 4, measured at symbols) were evoked by uncaging in the astrocytes marked 1–4. Time points for **k–n** are shown in series 1. Asterisk in **k** indicates the endfoot measured for series 1. **p**, Summary showing astrocyte endfeet fluorescence when constriction was observed in comparison with when there was no constriction. Results are shown as means  $\pm$  s.e.m. **q**, Infrared-transmitted and Fluo-3 fluorescence image (left) showing two dye-coupled astrocytes that contact the vasculature with their endfeet; whole-cell experiment summary (right) showing  $\Delta F/F$  when constriction occurred in comparison with no constriction. Results are shown as means  $\pm$  s.e.m. Scale bars, 20  $\mu m$  (micrographs); 50%  $\Delta F/F$ , 1 s (**e**); 100%  $\Delta F/F$ , 60 s (**j**, **o**); 20% constriction (**o**). **p**, **q**, Numbers in parentheses represent  $n$  values for endfeet above the bars and arterioles below the bars.



microscopy<sup>7,8</sup>. Brain slices were obtained from both rats and transgenic mice in which the astrocytes were labelled with green fluorescent protein (GFP) driven by the promoter for glial fibrillary acidic protein (GFAP)<sup>9</sup>. Astrocytes and their endfeet were selectively loaded with the acetoxymethyl (AM) ester form of the  $\text{Ca}^{2+}$  indicator Rhod-2. Neurons and vascular elements did not load with indicator in any imaging experiments ( $n = 213$  brain slices) (Fig. 1a–d; Supplementary movie 1). We co-loaded slices with the AM ester  $\text{Ca}^{2+}$  cage DMNP-EDTA to increase  $[\text{Ca}^{2+}]_i$  selectively in astrocytes (Fig. 1e). The initial increase triggered  $\text{Ca}^{2+}$  waves that propagated through the astrocyte syncytium ( $6.3 \pm 1.7 \mu\text{m s}^{-1}$ ,  $n = 39$  brain slices) (Fig. 1f–j; Supplementary movie 2). In control experiments ( $n = 17$  brain slices), astrocytes were not co-loaded with the  $\text{Ca}^{2+}$  cage DMNP-EDTA: increases in  $\text{Ca}^{2+}$  were observed in astrocytes only at extreme laser intensities that caused non-recovering fluorescent transients and morphological damage to the astrocytes ( $n = 68$  astrocytes). These increases in  $\text{Ca}^{2+}$  were isolated to the intensely illuminated astrocytes (data not shown).

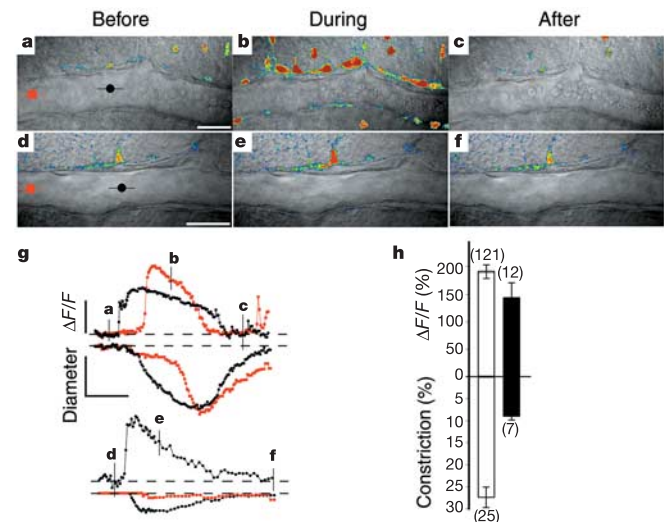
The selective induction of  $\text{Ca}^{2+}$  waves in astrocytes allowed us to determine how the diameters of cerebral blood vessels were altered when  $\text{Ca}^{2+}$  waves propagated into the endfeet of astrocytes at the vessel wall. When  $\text{Ca}^{2+}$  waves reached the astrocyte endfeet and induced large increases in  $[\text{Ca}^{2+}]_i$  ( $\Delta F/F = 185.6 \pm 9.9\%$ ,  $n = 133$  astrocytes in 32 brain slices), a constriction was observed in the adjacent region of the arteriole ( $23.4 \pm 2.2\%$ ,  $n = 32$  brain slices) (Fig. 1k–n; Supplementary movie 3). The peak increase in  $[\text{Ca}^{2+}]_i$  in the endfeet preceded the initial onset of vessel constriction by  $2.7 \pm 0.5$  s ( $n = 32$  brain slices). Constrictions could be induced repetitively in the same arteriole with repeated  $\text{Ca}^{2+}$  waves (Fig. 1o; Supplementary movie 4). Vascular constrictions ( $30.8 \pm 5.1\%$ ,  $n = 3$ ) still occurred when tetrodotoxin was added to block neuronal activity. Dilation was never observed and intense laser illumination of neurons during these experiments did not induce either  $\text{Ca}^{2+}$  waves or constrictions. When  $\text{Ca}^{2+}$  waves produced changes in  $[\text{Ca}^{2+}]_i$  only in the soma regions of the astrocytes without propagation into the endfeet or only small  $[\text{Ca}^{2+}]_i$  changes were observed in the endfeet ( $\Delta F/F = 82.5 \pm 1.24\%$ ;  $n = 112$  astrocytes in 28 brain slices), no constriction occurred ( $-0.22 \pm 0.63\%$ ,  $n = 28$  brain slices) (Fig. 1p).

Whole-cell patch-clamp experiments were performed in astrocytes identified by morphological criteria with the use of infrared differential interference contrast (IR-DIC) optics<sup>10</sup> and electrophysiological properties<sup>11</sup> (RMP;  $-84.0 \pm 1.4$  mV ( $n = 27$ )). Astrocytes were loaded with the cell-membrane-impermeant form of the  $\text{Ca}^{2+}$  cage DMNP-EDTA and the low-molecular-mass form of Fluo-3 (854 Da), that allowed for limited dye-coupling between astrocytes and fluorescence imaging in up to three astrocyte endfeet at the vessel wall (Fig. 1q). Confirming the observations above, when two-photon photolysis of the  $\text{Ca}^{2+}$  cage induced  $\text{Ca}^{2+}$  waves that propagated into the endfeet of astrocytes and induced large increases in  $[\text{Ca}^{2+}]_i$  ( $\Delta F/F = 196.5 \pm 28.8\%$ ;  $n = 26$  astrocytes, 17 brain slices), a constriction was observed in the adjacent region of the arteriole ( $18.6 \pm 2.4\%$ ,  $n = 17$  brain slices; Fig. 1q). Dilation was never observed and intense laser illumination of neurons during these experiments did not induce either  $\text{Ca}^{2+}$  waves or constrictions. No constriction occurred ( $-0.78 \pm 0.96\%$ ,  $n = 10$  brain slices; Fig. 1q) when  $\text{Ca}^{2+}$  waves produced changes in  $[\text{Ca}^{2+}]_i$  only in the soma regions of the astrocytes without propagation into the endfeet or when small  $[\text{Ca}^{2+}]_i$  changes were observed in the endfeet ( $\Delta F/F = 79.0 \pm 16\%$ ;  $n = 20$  astrocytes, 10 brain slices).

We examined the influence of increases in endfeet  $[\text{Ca}^{2+}]_i$  when the increase was widespread and involved many endfeet in comparison with when the  $\text{Ca}^{2+}$  wave propagated to only one or two endfeet. If the  $\text{Ca}^{2+}$  wave propagated along the vascular profile and caused an increase in numerous astrocyte endfeet then the constriction was equally extensive (Fig. 2a–c, g). However, if the  $\text{Ca}^{2+}$  wave propagated to only one or two endfeet and did not propagate

extensively along the arteriole, the adjacent arteriole constriction was also similarly restricted and did not propagate away from this site (Fig. 2d–g; Supplementary movie 5). The difference in the magnitude of arteriole constriction was considerable between experiments involving multiple endfeet and experiments when the increase in  $[\text{Ca}^{2+}]_i$  was restricted to one or two endfeet. Arterioles constricted by  $27.5 \pm 2.2\%$  ( $n = 25$  brain slices) when multiple astrocyte endfeet controlled constriction compared with  $9.1 \pm 0.75\%$  ( $n = 7$  brain slices) when only one or two endfeet were involved (Fig. 2h). The magnitude of the endfeet  $[\text{Ca}^{2+}]_i$  changes did not differ significantly:  $\Delta F/F = 192.1 \pm 10.8$  (121 endfeet) and  $\Delta F/F = 145.0 \pm 26.3$  (12 endfeet), respectively.

The cerebral microvasculature is densely innervated by noradrenaline projections from the locus coeruleus region that have been shown to synapse principally on astrocyte endfeet<sup>4,5</sup>. Activation by noradrenaline *in vivo* produces vasoconstriction and reduced CBF<sup>2,3</sup>. Noradrenaline also increases  $[\text{Ca}^{2+}]_i$  in astrocytes<sup>12</sup>. Application of noradrenaline (10  $\mu\text{M}$ ) in a bath evoked  $[\text{Ca}^{2+}]_i$  increases in astrocyte endfeet ( $\Delta F/F = 178.8 \pm 6.2$ ,  $n = 225$  endfeet at 61 vessels) and reproducible arteriole constrictions ( $29.1 \pm 1.4\%$ ,  $n = 61$ ) (Supplementary Fig. 1). Arteriole constrictions never occurred without large-amplitude  $[\text{Ca}^{2+}]_i$  increases in astrocyte endfeet. We examined the link between  $[\text{Ca}^{2+}]_i$  increases in astrocyte endfeet and vasoconstrictions induced by noradrenaline. Local application of noradrenaline (10  $\mu\text{M}$ ) increased  $[\text{Ca}^{2+}]_i$  in the astrocyte endfeet ( $\Delta F/F = 157.4 \pm 7.4$ ,  $n = 31$  endfeet at 14 vessels) and produced arteriole constrictions ( $26.7 \pm 1.7\%$ ,  $n = 14$ ) (Fig. 3). The peak astrocyte endfeet  $[\text{Ca}^{2+}]_i$  preceded the onset of arteriole constriction by  $1.24 \pm 0.22$  s (Fig. 3). To evaluate the effect of noradrenaline in the absence of astrocytic endfeet,  $\text{Ca}^{2+}$  signalling astrocytes were loaded with the AM form of the high-affinity  $\text{Ca}^{2+}$  chelator BAPTA. Application of noradrenaline (10  $\mu\text{M}$ ) in a bath in slices loaded with BAPTA-AM (100  $\mu\text{M}$ ) evoked minimal



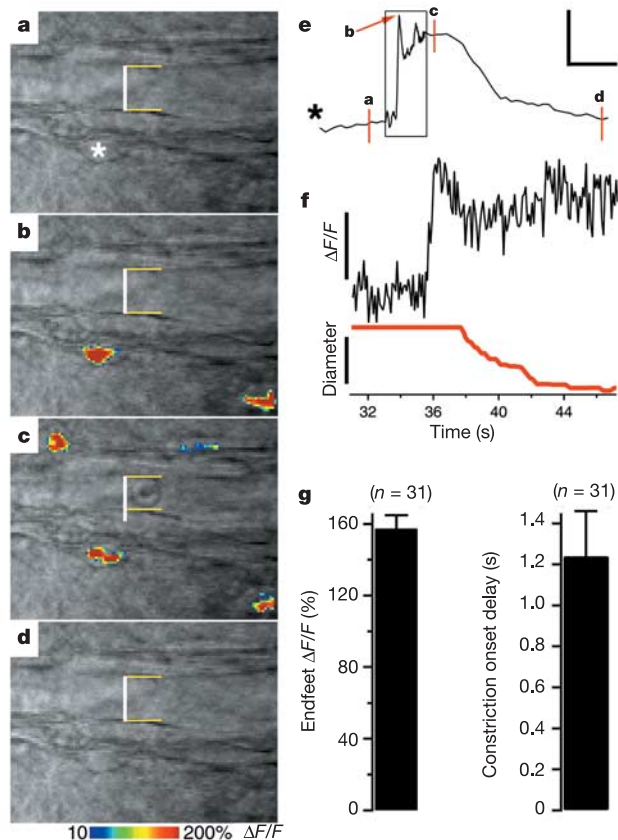
**Figure 2** The extent of vascular constrictions was related to the number of endfeet showing increased  $[\text{Ca}^{2+}]_i$ . Vascular constrictions were greater and more widespread when  $[\text{Ca}^{2+}]_i$  increased in numerous endfeet (a–c) compared with only one endfeet (d–f). a–g, IR-transmitted and Rhod-2 fluorescence (pseudocolour) overlay images (a–f) and measurements (g, taken at the symbols) showing two different vessels and their associated endfeet  $[\text{Ca}^{2+}]_i$  before, during and after maximal increases in endfeet  $[\text{Ca}^{2+}]_i$  resulting from  $\text{Ca}^{2+}$  waves. Scale bars a–f, 30  $\mu\text{m}$ . Image time points (a–f) are marked on the  $\text{Ca}^{2+}$  transients. Scale bars g, 100%  $\Delta F/F$ , 20% diameter change, 60 s. h,  $[\text{Ca}^{2+}]_i$  increases in multiple endfeet (non-filled bars) in comparison with one or two endfeet (filled bars) and associated constrictions. Results are shown as means  $\pm$  s.e.m. Numbers in parentheses represent  $n$  values for endfeet above the bars and arterioles below the bars.



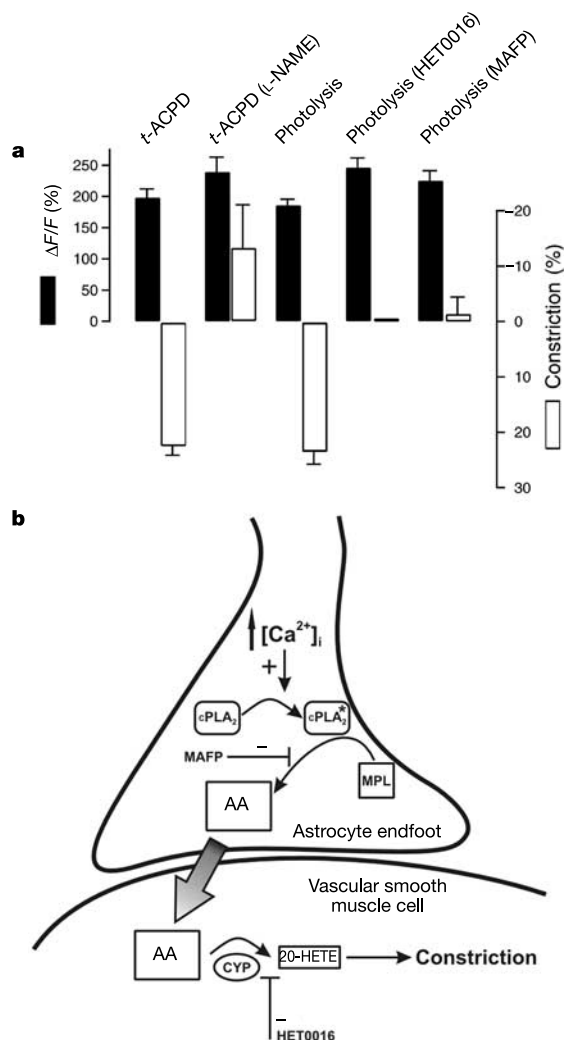
$\text{Ca}^{2+}$  increases in astrocyte endfeet ( $\Delta F/F = 19.8 \pm 6.4$ ,  $n = 55$  endfeet at nine vessels) and small arteriole constrictions ( $9.9 \pm 2.5\%$ ,  $n = 9$ ) (Supplementary Fig. 2). It is unlikely that BAPTA was in smooth muscle cells because we never saw any indication of loading in smooth muscle cells with AM dyes, and direct electrical stimulation of arterioles in these experiments could still induce constriction. It is possible that noradrenaline might have additional direct actions on the vessels themselves.

The time delay of several seconds between the increase in  $[\text{Ca}^{2+}]_i$  in the endfeet and the vascular constriction suggests the formation and release of an intercellular second messenger that diffuses from the astrocyte endfoot to the vascular cells. We tested the involvement of the  $\text{Ca}^{2+}$ -sensitive phospholipase  $\text{A}_2$  ( $\text{PLA}_2$ ), which is principally expressed in astrocytes<sup>13</sup> and causes the release of arachidonic acid from increased  $[\text{Ca}^{2+}]_i$ <sup>14</sup>. Blocking  $\text{PLA}_2$  with 100  $\mu\text{M}$  methyl arachidonyl fluorophosphonate (MAFP)<sup>15</sup> eliminated vascular constrictions ( $-1.3 \pm 3.1\%$ ,  $n = 6$ ; Fig. 4a). The major vasoconstrictive metabolite 20-hydroxyeicosatetraenoic acid (20-HETE)<sup>16</sup> is generated copiously in cerebrovascular smooth muscle cells by CYP4A, a cytochrome P450 (CYP450) enzyme subtype, in response to elevated arachidonic acid concentrations<sup>17</sup>. 20-HETE blocks calcium-activated  $\text{K}^+$  channels in smooth muscle cells, leading to depolarization, increased  $\text{Ca}^{2+}$  influx and contraction<sup>16</sup>. Inhibiting 20-HETE formation with the selective CYP4A

inhibitor HET0016 (ref. 18) (100  $\mu\text{M}$ ) blocked vascular constrictions ( $0.1 \pm 0.1\%$ ,  $n = 8$ ; Fig. 4a). A recent study<sup>19</sup> indicated that  $\text{Ca}^{2+}$  increases in astrocytes might induce vasodilation. However, in that study most slices were preincubated in  $N^G$ -nitro-L-arginine methyl ester (L-NAME) to block NO formation and precontract blood vessels<sup>19</sup>. In contrast, we found that in untreated control slices the mGluR agonist 1-aminocyclopentane-*trans*-1,3-dicarboxylic acid (*t*-ACPD; 100  $\mu\text{M}$ ), increased  $\text{Ca}^{2+}$  in astrocyte endfeet and induced robust vascular constrictions ( $22.3 \pm 1.6\%$ ,  $n = 6$ , Fig. 4a). Incubating slices in L-NAME (100  $\mu\text{M}$ ) abolished the *t*-ACPD induced constrictions that we observed in control conditions and induced a small dilation ( $-13.2 \pm 7.8\%$ ,  $n = 6$ , Fig. 4a). We therefore propose that, under physiological conditions, increased  $[\text{Ca}^{2+}]_i$  in astrocyte endfeet activates  $\text{PLA}_2$ , inducing the formation and release of arachidonic acid, which is converted to 20-HETE by CYP450 enzymes in smooth muscle cells to induce contraction



**Figure 3** Noradrenaline-induced  $\text{Ca}^{2+}$  elevations in astrocyte endfeet precede the onset of vessel constriction. **a–d**, Overlay images (**a–d**) showing the arteriole and endfeet  $[\text{Ca}^{2+}]_i$  responses to noradrenaline (10  $\mu\text{M}$ ). Peak astrocyte endfoot response (asterisk, timecourse in **e** with frame times) precedes the onset of arteriole constriction. The control diameter was 11  $\mu\text{m}$ . **f**, Expanded timecourse from **e** (boxed area) showing endfoot  $[\text{Ca}^{2+}]_i$  increases before vessel constriction. Scale bars, 20 s for **e**, 100%  $\Delta F/F$  for **e** and **f**, 20% diameter change for **f**. **g**, Average amplitude of endfeet  $[\text{Ca}^{2+}]_i$  in noradrenaline and the time between the peak astrocyte endfeet  $[\text{Ca}^{2+}]_i$  and arteriole constriction onset. Results are shown as means  $\pm$  s.e.m.



**Figure 4** Increased  $[\text{Ca}^{2+}]_i$  in astrocyte endfeet causes cerebrovascular constrictions. **a**, Summary showing Rhod-2  $\Delta F/F$  in endfeet and associated constriction under several conditions. mGluR receptor activation by *t*-ACPD is compared in control versus with L-NAME to block NO generation.  $\text{Ca}^{2+}$  photolysis is compared in control conditions versus with CYP4A inhibition (HET0016), or with phospholipase  $\text{A}_2$  inhibition (MAFP). Results are shown as means  $\pm$  s.e.m. **b**, The proposed mechanism of astrocyte endfeet  $\text{Ca}^{2+}$ -induced vascular constrictions. Increased  $[\text{Ca}^{2+}]_i$  in the astrocyte endfeet activates cytosolic  $\text{Ca}^{2+}$ -activated  $\text{PLA}_2$  and increases arachidonic acid (AA) formation from membrane phospholipids (MPL). AA freely diffuses to smooth muscle cells to be converted to 20-HETE by CYP4A and thus to induce vascular constrictions.

(Fig. 4b). CYP450 enzymes and 20-HETE formation are inhibited by nitric oxide<sup>16</sup>. Generation of nitric oxide by neuronal activity might contribute to vasodilation by preventing the vasoconstriction induced by arachidonic acid released from astrocytes.

We conclude that astrocyte  $\text{Ca}^{2+}$  waves cause vascular constrictions when the endfeet participate in the  $\text{Ca}^{2+}$  wave. This represents a previously unknown mechanism of CBF regulation. Vasoactive neuronal projections that control CBF and extend the upper range of autoregulation<sup>20</sup> might mediate these actions through  $\text{Ca}^{2+}$ -dependent processes in astrocyte endfeet. Finally, these observations provide a link between increased  $[\text{Ca}^{2+}]_i$  in astrocytes and haemodynamic pathology. Increased  $[\text{Ca}^{2+}]_i$  in astrocytes during spreading depression<sup>21</sup> might cause the associated phase of vasoconstriction<sup>22</sup>. During a stroke, the ischaemic insult increases astrocyte  $[\text{Ca}^{2+}]_i$  (ref. 23), which might induce vascular constrictions to decrease CBF further and exacerbate infarct damage. Astrocyte  $[\text{Ca}^{2+}]_i$  signalling might therefore be a novel therapeutic target for treatment of stroke and migraine. □

## Methods

### Slice preparation and AM-ester loading

Brain slices were obtained from both rats and GFAP/GFP transgenic mice (strain FVB/N-TgN(GFAPGFP)14Mes; Jackson Labs)<sup>9</sup> aged 13–18 days postnatal, as described previously<sup>1</sup>. Slices were stored at room temperature (20 to 22°C) for 1 h before the loading of dye/ $\text{Ca}^{2+}$  chelator in an oxygenated artificial cerebrospinal fluid (aCSF) containing (in mM): NaCl 126, KCl 2.5 or 4.2,  $\text{NaHCO}_3$  26, glucose 10,  $\text{MgCl}_2$  2,  $\text{NaH}_2\text{PO}_4$  1.25 and  $\text{CaCl}_2$  2. Astrocytes were co-loaded with the  $\text{Ca}^{2+}$  indicator Rhod-2-AM (10  $\mu\text{M}$ ) and the  $\text{Ca}^{2+}$  cage DMNP-EDTA-AM (10  $\mu\text{M}$ ) (Molecular Probes, Eugene, Oregon) for 1 h (DMSO concentration 0.3%). For the  $\text{Ca}^{2+}$  chelation/noradrenaline experiments, the astrocytes were co-loaded with the  $\text{Ca}^{2+}$  indicator Rhod-2-AM (10  $\mu\text{M}$ ) and the membrane-permeable (AM ester) form of the  $\text{Ca}^{2+}$  chelator BAPTA (100  $\mu\text{M}$ ) (Molecular Probes) for more than 1 h (DMSO concentration 0.3%) and the multidrug resistance protein transport inhibitor probenidol (1–3 mM; Sigma, Oakville, Ontario). Slices were transferred to a recording chamber and perfused with oxygenated aCSF at a rate of 1–3 ml min<sup>-1</sup> and maintained at either 25 ± 0.5 or 33 ± 0.5°C with an inline heater (Warner Instruments). In some experiments ( $n = 18$ ) the extracellular  $[\text{K}^+]$  was increased to 4.2 mM to ensure that vascular responses were not altered with higher  $[\text{K}^+]$ .

### Two-photon imaging and photolysis

We performed imaging with a two-photon laser-scanning microscope (Zeiss LSM510-Axiokop-2 fitted with a 40X-W/0.80 numerical aperture objective lens) directly coupled to a Mira Ti:sapphire laser (~100-fs pulses, 76 MHz, pumped by a 5 W Verdi laser; Coherent). GFP and the Rhod-2 fluorophore, which have large two-photon absorption cross-sections<sup>24</sup>, were excited at 835–840 nm and epifluorescence was detected with external detectors with 510-nm (40 nm bandpass) and 605-nm (55 nm bandpass) filters. For two-photon photolysis within single or multiple astrocytes we tuned the Ti:sapphire laser to 730 nm for maximum  $\text{Ca}^{2+}$  uncaging of DMNP-EDTA<sup>8</sup> and excitation of Rhod-2. Laser intensities were the lowest possible for uncaging. Flash durations between 3 and 10 ms within single astrocytes produced repeatable small transients. The laser intensity was carefully increased beyond the point for the initial transient event until a large  $\text{Ca}^{2+}$  transient characteristic of internal release occurred within the astrocyte that then induced a  $\text{Ca}^{2+}$  wave that propagated throughout the astrocyte network. Because of the highly nonlinear nature of two-photon photolysis of caged  $\text{Ca}^{2+}$  (ref. 8), no uncaging occurs during Rhod-2 fluorescence acquisition, in which laser intensity was typically one-tenth of the intensity necessary for photolysis.

### Whole-cell patch recording

Astrocytes, identified using IR-DIC optics, were recorded in whole-cell current-clamp mode with an Axopatch 200B amplifier. Pipette (3–7 M $\Omega$ ) solution (pH 7.2–7.4) contained (in mM): potassium gluconate 130, KCl 10, HEPES 10, NaATP 4, TrisGTP 0.3, Fluo-3 pentaammonium salt (854 Da; Molecular Probes) 0.1–0.3, and DMNP-EDTA 2 (the cell-membrane-impermeant form of the  $\text{Ca}^{2+}$  cage; Molecular Probes) loaded with  $\text{Ca}^{2+}$  (60–75%). Fluo-3 fluorescence was detected with a 510-nm (40 nm bandpass) emission filter and 488-laser line excitation (Zeiss LSM510).

### Analysis of $\text{Ca}^{2+}$ signals and arteriole diameter

Arterioles (lumen diameter 15.4 ± 0.7  $\mu\text{m}$ ) were identified with IR-DIC optics and chosen on the basis of their diameter, appearance of smooth muscle cells and healthy appearance over an extended region. Although blood vessels in brain slices do not experience the same shear stresses as they do *in vivo*, repeatable constrictions could be induced. Vessel diameter changes were imaged by acquiring 730-nm-wavelength laser transmission with an external photomultiplier simultaneously with the Rhod-2 and Fluo-3 fluorescence. Image series were analysed off-line, and luminal diameter measurements were made at multiple sites along the vessel and quantified with Zeiss LSM (version 2.8) software and custom-made autotracking software (Diamtrak, developed by

T. O. Neild). Fluorescence signals were defined as  $\Delta F/F = [(F_1 - B_1) - (F_0 - B_0)] / (F_0 - B_0)$ , where  $F_1$  and  $F_0$  are fluorescence in the astrocyte endfoot at any given time point and at the beginning of the experiment respectively, and  $B_1$  and  $B_0$  are the background fluorescence at the same time point and at the beginning of the experiment respectively. Background values were taken from an adjacent area located at least 10  $\mu\text{m}$  from imaged areas. Results are shown as means ± s.e.m.

### Drug application experiments

*N*-Hydroxy-*N'*-(4-butyl-2-methylphenyl)-formamidine (HET0016, a gift from Taisho Pharmaceutical Co.) the inhibitor of 20-HETE, and MAFP, the inhibitor of PLA<sub>2</sub>, were applied by incubating brain slices for 1–2 h before transfer to the perfusion and imaging chamber. *t*-ACPD (100  $\mu\text{M}$ ) and noradrenaline (10  $\mu\text{M}$  in aCSF) were applied in a bath. Noradrenaline was also locally applied through a manipulator-controlled pipette (~2 M $\Omega$ ) by a picospritzer with less than 1 lb in<sup>-2</sup> air pressure. Noradrenaline was prepared freshly before each application. Arteriole transmitted and astrocyte endfoot fluorescence images were acquired at 85–147-ms frame rates. The delay in constriction onset was defined as the difference between the initial time point of peak astrocyte endfoot fluorescence and the initial arteriole constriction event.

Received 23 April; accepted 9 July 2004; doi:10.1038/nature02827.

- Chaigneau, E., Oheim, M., Audinat, E. & Charpak, S. Two-photon imaging of capillary blood flow in olfactory bulb glomeruli. *Proc. Natl Acad. Sci. USA* **100**, 13081–13086 (2003).
- Raichle, M. E., Hartman, B. K., Eichling, J. O. & Sharpe, L. G. Central noradrenergic regulation of cerebral blood flow and vascular permeability. *Proc. Natl Acad. Sci. USA* **72**, 3726–3730 (1975).
- Goadsby, P. J. & Edvinsson, L. in *Cerebral Blood Flow and Metabolism* (ed. Krause, D. N.) 172–188 (Lippincott Williams & Wilkins, Philadelphia, 2002).
- Cohen, Z., Molinatti, G. & Hamel, E. Astroglial and vascular interactions of noradrenaline terminals in the rat cerebral cortex. *J. Cereb. Blood Flow Metab.* **17**, 894–904 (1997).
- Paspalas, C. D. & Papadopoulos, G. C. Ultrastructural relationships between noradrenergic nerve fibers and non-neuronal elements in the rat cerebral cortex. *Glia* **17**, 133–146 (1996).
- Simard, M., Arcuino, G., Takano, T., Liu, Q. S. & Nedergaard, M. Signaling at the gliovascular interface. *J. Neurosci.* **23**, 9254–9262 (2003).
- Soeller, C. & Cannell, M. B. Two-photon microscopy: imaging in scattering samples and three-dimensionally resolved flash photolysis. *Microsc. Res. Tech.* **47**, 182–195 (1999).
- Brown, E. B., Shear, J. B., Adams, S. R., Tsien, R. Y. & Webb, W. W. Photolysis of caged calcium in femtoliter volumes using two-photon excitation. *Biophys. J.* **76**, 489–499 (1999).
- Zhuo, L. *et al.* Live astrocytes visualized by green fluorescent protein in transgenic mice. *Dev. Biol.* **187**, 36–42 (1997).
- Kang, J. & Nedergaard, M. in *Imaging Neurons: A Laboratory Manual* (ed. Konnerth, A.) 42.1–42.11 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2000).
- Walz, W. & MacVicar, B. Electrophysiological properties of glial cells: comparison of brain slices with primary cultures. *Brain Res.* **443**, 321–324 (1988).
- Duffy, S. & MacVicar, B. A. Adrenergic calcium signaling in astrocyte networks within the hippocampal slice. *J. Neurosci.* **15**, 5535–5550 (1995).
- Farooqui, A. A., Yang, H. C., Rosenberger, T. A. & Horrocks, L. A. Phospholipase A<sub>2</sub> and its role in brain tissue. *J. Neurochem.* **69**, 889–901 (1997).
- Katsuki, H. & Okuda, S. Arachidonic acid as a neurotoxic and neurotrophic substance. *Prog. Neurobiol.* **46**, 607–636 (1995).
- Street, I. P. *et al.* Slow- and tight-binding inhibitors of the 85-kDa human phospholipase A<sub>2</sub>. *Biochemistry* **32**, 5935–5940 (1993).
- Roman, R. J. P-450 metabolites of arachidonic acid in the control of cardiovascular function. *Physiol. Rev.* **82**, 131–185 (2002).
- Gebremedhin, D. *et al.* Production of 20-HETE and its role in autoregulation of cerebral blood flow. *Circ. Res.* **87**, 60–65 (2000).
- Miyata, N. *et al.* HET0016, a potent and selective inhibitor of 20-HETE synthesizing enzyme. *Br. J. Pharmacol.* **133**, 325–329 (2001).
- Zonta, M. *et al.* Neuron-to-astrocyte signaling is central to the dynamic control of brain microcirculation. *Nature Neurosci.* **6**, 43–50 (2003).
- Chillon, J. M. & Baumbach, G. L. in *Cerebral Blood Flow and Metabolism* (ed. Krause, D. N.) 395–412 (Lippincott Williams & Wilkins, Philadelphia, 2002).
- Basarsky, T. A., Duffy, S. N., Andrew, R. D. & MacVicar, B. A. Imaging spreading depression and associated intracellular calcium waves in brain slices. *J. Neurosci.* **18**, 7189–7199 (1998).
- Dreier, J. P. *et al.* Ischaemia triggered by spreading neuronal activation is inhibited by vasodilators in rats. *J. Physiol. (Lond.)* **531**, 515–526 (2001).
- Duffy, S. & MacVicar, B. A. *In vitro* ischemia promotes calcium influx and intracellular calcium release in hippocampal astrocytes. *J. Neurosci.* **16**, 71–81 (1996).
- Xu, C., Zipfel, W., Shear, J. B., Williams, R. M. & Webb, W. W. Multiphoton fluorescence excitation: new spectral windows for biological nonlinear microscopy. *Proc. Natl Acad. Sci. USA* **93**, 10763–10768 (1996).

**Supplementary Information** accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank A. G. Phillips, Y. T. Wang and T. Murphy for comments on the manuscript, and D. Feighan for technical assistance. S.J.M. was supported by a Canadian Heart and Stroke Fellowship. B.A.M. is a Canada Research Chair in Neuroscience and Michael Smith Distinguished Scholar. Work was supported by Canadian Institutes of Health Research and the Canadian Stroke Network.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to B.A.M. (bmavica@interchange.ubc.ca).

# Absence of S6K1 protects against age- and diet-induced obesity while enhancing insulin sensitivity

Sung Hee Um<sup>1</sup>, Francesca Frigerio<sup>1</sup>, Mitsuhiro Watanabe<sup>2</sup>, Frédéric Picard<sup>2\*</sup>, Manel Joaquin<sup>1</sup>, Melanie Sticker<sup>1</sup>, Stefano Fumagalli<sup>1</sup>, Peter R. Allegrini<sup>3</sup>, Sara C. Kozma<sup>1\*</sup>, Johan Auwerx<sup>2</sup> & George Thomas<sup>1</sup>

<sup>1</sup>Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, 4058 Basel, Switzerland

<sup>2</sup>Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM/ULP, and Institut Clinique de la Souris, Génomole Strasbourg, 67404 Illkirch, France

<sup>3</sup>Novartis Pharma AG, Klybeckstrasse 141, 4057 Basel, Switzerland

\* Present address: Laval Hospital Research Center, 2725 chemin Ste-Foy, Ste-Foy, Quebec G1V 4G5, Canada (F.P.); Genome Research Institute, University of Cincinnati, 2180 E. Galbraith Road, Cincinnati, Ohio 45237, USA (S.C.K.)

Elucidating the signalling mechanisms by which obesity leads to impaired insulin action is critical in the development of therapeutic strategies for the treatment of diabetes<sup>1</sup>. Recently, mice deficient for S6 Kinase 1 (S6K1), an effector of the mammalian target of rapamycin (mTOR) that acts to integrate nutrient and insulin signals<sup>2</sup>, were shown to be hypoinsulinaemic, glucose intolerant and have reduced  $\beta$ -cell mass<sup>3</sup>. However, S6K1-deficient mice maintain normal glucose levels during fasting, suggesting hypersensitivity to insulin<sup>3</sup>, raising the question of their metabolic fate as a function of age and diet. Here, we report that S6K1-deficient mice are protected against obesity owing to enhanced  $\beta$ -oxidation. However on a high fat diet, levels of glucose and free fatty acids still rise in S6K1-deficient mice, resulting in insulin receptor desensitization. Nevertheless, S6K1-deficient mice remain sensitive to insulin owing to the apparent loss of a negative feedback loop from S6K1 to insulin receptor substrate 1 (IRS1), which blunts S307 and S636/S639 phosphorylation; sites involved in insulin resistance<sup>4,5</sup>. Moreover, wild-type mice on a high fat diet as well as *K/K A'* and *ob/ob* (also known as *Lep/Lep*) mice—two genetic models of obesity—have markedly elevated S6K1 activity and, unlike S6K1-deficient mice, increased phosphorylation of IRS1 S307 and S636/S639. Thus under conditions of nutrient satiation S6K1 negatively regulates insulin signalling.

As animals reach adulthood their growth rate decreases and fatty acids are largely converted into triglycerides and stored as an energy reserve in adipose tissue. To investigate the effect of age on growth, a matched set of S6K1-deficient (*S6K1*<sup>-/-</sup>) and wild-type male mice were placed on a normal chow diet (NCD) (4% total calories derived from fat, 3,035 kcal kg<sup>-1</sup>) and monitored over a period of 17 weeks from 10 weeks of age. The rate at which *S6K1*<sup>-/-</sup> mice increased body weight on the NCD was significantly reduced compared with wild-type mice: at 27 weeks of age the difference in body weight was 25% (Fig. 1a). Dissection of *S6K1*<sup>-/-</sup> mice revealed a marked reduction in epididymal white adipose tissue (WAT) (Supplementary Fig. 1a). When normalized for body weight, epididymal, inguinal and retroperitoneal fat pads (Fig. 1b), as well as the brown fat pad (Supplementary Fig. 1b) were significantly reduced. Furthermore, the decrease was specific for fat, as the weight of other major organs such as liver was not affected after correction for total body weight (Fig. 1b).

Analysis of adipocytes in epididymal fat pads by either scanning electron microscopy or by haematoxylin and eosin staining showed a sharp reduction in size, with some adipocytes exhibiting a multi-locular-like phenotype (Fig. 1c and see below). Morphometric analysis revealed that *S6K1*<sup>-/-</sup> adipocytes were consistently smaller

compared with adipocytes from wild-type mice (Fig. 1c), with an average 71% decrease in size (Supplementary Fig. 1c). The decrease in fat accumulation in *S6K1*<sup>-/-</sup> mice was not due to less food intake, which was increased 17% when food consumption was adjusted for body weight (Fig. 1d). Moreover, on the basis of normal fasting and feeding glucose levels and no increase in ketone body formation<sup>3</sup> (Table 1), *S6K1*<sup>-/-</sup> mice did not seem to be starving, nor was there an alteration in adaptive thermogenesis (data not shown). This raised the possibility that triglycerides were being broken down rather than stored in WAT. Consistent with this, basal rates of lipolysis were fivefold higher in *S6K1*<sup>-/-</sup> versus wild-type adipocytes, although norepinephrine-induced fatty acid and glycerol release increased in both genotypes in a dose-dependent manner to the same final extent (Fig. 1e; see also Supplementary Fig. 1d). Moreover the metabolic rate was greatly enhanced in *S6K1*<sup>-/-</sup> mice, as indicated by the 27% increase in oxygen consumption versus wild-type mice (Fig. 1f). The respiratory exchange ratio (RER) of 0.713  $\pm$  0.004 for wild-type mice and 0.709  $\pm$  0.003 for *S6K1*<sup>-/-</sup> mice ( $P < 0.01$ ) showed that both animals were largely using fatty acids as an energy source. Thus the failure to accumulate fat with age in *S6K1*<sup>-/-</sup> mice seems to stem from a sharp increase in lipolytic and metabolic rates.

These increased responses, combined with the finding that levels of circulating triglycerides and free fatty acids (FFAs) were similar in both genotypes (Table 1), suggested that in *S6K1*<sup>-/-</sup> mice triglycerides were being rapidly oxidized in WAT and/or muscle. WAT is not an energy-consuming tissue; however, electron micrographs revealed multi-locular adipocytes, with mitochondria of increased size and number—phenotypes that are absent in wild-type adipocytes (Fig. 2a). Consistent with this, analysis of the messenger RNA levels of genes involved in energy combustion and oxidative phosphorylation were found to be strongly increased in *S6K1*<sup>-/-</sup> adipocytes compared with wild-type adipocytes, including uncoupling protein 1 (UCP1), UCP3, carnitine palmitoyltransferase 1 (CPT1) and PPAR- $\gamma$  co-activator 1- $\alpha$  (PGC1- $\alpha$ ) (Fig. 2a; see also Supplementary Fig. 2a). Mitochondrial content was also affected in *S6K1*<sup>-/-</sup> skeletal muscle (Fig. 2b), consistent with increased expression of peroxisome proliferator-activated receptor  $\beta/\delta$  (PPAR- $\beta/\delta$ ), PGC1- $\alpha$ , UCP3 and CPT1 (Fig. 2b; see also Supplementary Fig. 2b). As *S6K1*<sup>-/-</sup> mice have reduced WAT and increased oxidative phosphorylation, this raised the possibility that *S6K1*<sup>-/-</sup> mice are protected against diet-induced obesity, which is linked to the oxidative phosphorylation pathway<sup>6–10</sup>. Indeed, when *S6K1*<sup>-/-</sup> mice were challenged with a high fat diet (HFD) (60% total calories derived from fat, 4,057 kcal kg<sup>-1</sup>) weight accumulation was significantly reduced compared with wild-type mice during the 4-month feeding period (Figs 2c, d). Although food intake in *S6K1*<sup>-/-</sup> mice is similar to that of wild-type mice, when normalized to body weight they consume 44% more food (Supplementary Fig. 2c). Metabolic rate measured by indirect calorimetry increased for both genotypes on the HFD, with the effect more pronounced for *S6K1*<sup>-/-</sup> mice (compare Figs 2e and 1f).

Table 1 One-hour post-prandial values after overnight fasting of wild-type and *S6K1*<sup>-/-</sup> mice

| Diet                                             | Wild type       |                 | <i>S6K1</i> <sup>-/-</sup> |                  |
|--------------------------------------------------|-----------------|-----------------|----------------------------|------------------|
|                                                  | Normal          | High fat        | Normal                     | High fat         |
| Insulin ( $\mu$ g l <sup>-1</sup> )              | 0.57 $\pm$ 0.09 | 2.22 $\pm$ 0.87 | 0.38 $\pm$ 0.08†           | 0.27 $\pm$ 0.06† |
| Triglycerides (mmol l <sup>-1</sup> )            | 0.61 $\pm$ 0.09 | 1.08 $\pm$ 0.15 | 0.55 $\pm$ 0.06            | 0.95 $\pm$ 0.14  |
| FFAs (mmol l <sup>-1</sup> )                     | 0.26 $\pm$ 0.05 | 0.27 $\pm$ 0.02 | 0.19 $\pm$ 0.04            | 0.56 $\pm$ 0.10† |
| Leptin (ng ml <sup>-1</sup> )                    | 6.11 $\pm$ 0.82 | 12.6 $\pm$ 0.00 | 3.22 $\pm$ 0.50            | 6.34 $\pm$ 2.35† |
| $\beta$ -Hydroxybutyrate (mg dl <sup>-1</sup> )* | 1.30 $\pm$ 0.05 | ND              | 1.40 $\pm$ 0.16            | ND               |

Values are for 6-month-old male mice of the indicated genotype fed a normal or high fat diet. Data represent mean  $\pm$  s.e.m. ND, not determined.

\* The level of  $\beta$ -hydroxybutyrate was measured in 6-h-fasted mice.

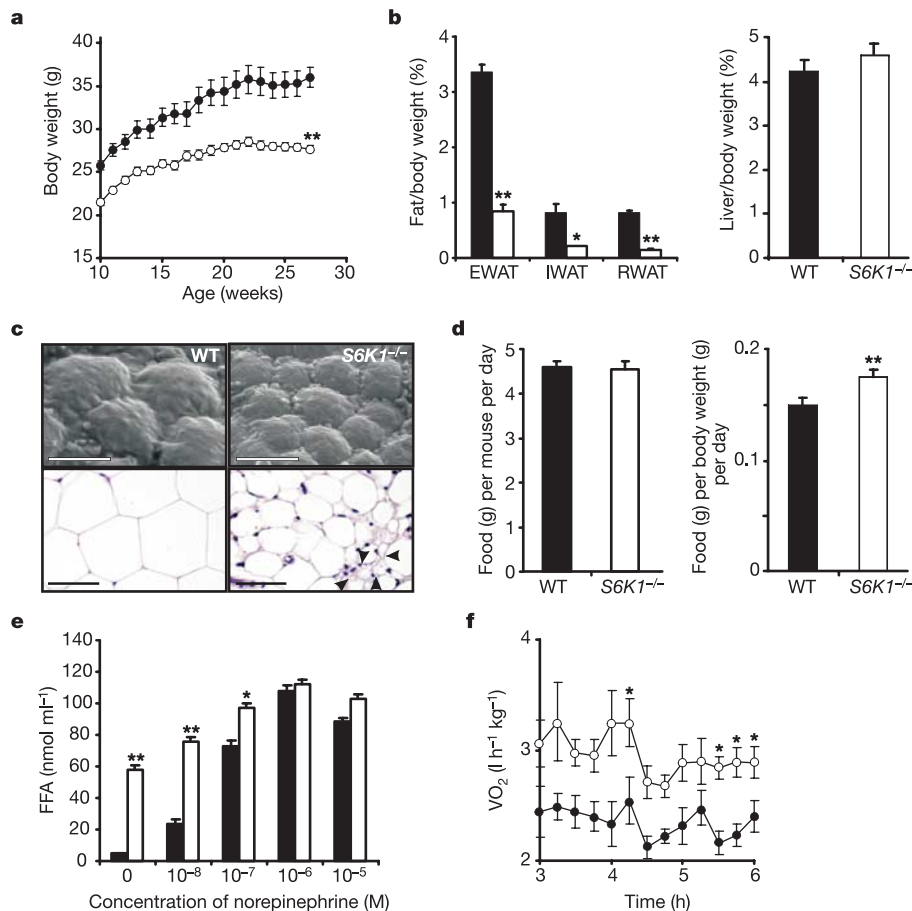
†  $P < 0.05$  compared with wild type ( $n = 6–18$ ).



Moreover RER remained unchanged in  $S6K1^{-/-}$  mice on a HFD ( $0.708 \pm 0.002$ ), whereas in wild-type mice it increased from  $0.713 \pm 0.004$  to  $0.729 \pm 0.002$  ( $n = 6$ ,  $P < 0.01$ ), indicating an increase in carbohydrate relative to fatty acid oxidation. To determine the weight gain corresponding to fat, mice were subjected to magnetic resonance imaging (MRI) analysis at 2-month intervals while on the HFD. A transverse section through the abdomen showed a marked reduction in fat depots, depicted by the less intense signal (Supplementary Fig. 2d). MRI assessment of total fat, corrected for body weight, revealed that the body fat index for  $S6K1^{-/-}$  mice increased by 20% over this period, whereas it doubled for wild-type mice (Fig. 2f). Thus, consistent with the increase in oxidative phosphorylation,  $S6K1^{-/-}$  mice are protected against diet-induced obesity.

Although  $S6K1^{-/-}$  mice have a high metabolic rate when on a HFD, they exhibit a threefold increase in circulating FFAs (Table 1), consistent with an increase in fat on their skin and hair (data not shown). As increased circulating FFAs are implicated in insulin resistance<sup>11–13</sup> and  $S6K1^{-/-}$  mice are hypoinsulinaemic<sup>3</sup>, it seemed likely that they would become insulin resistant on the HFD. On a NCD both  $S6K1^{-/-}$  and wild-type mice exhibited similar fasting levels of glucose (Fig. 3a), although  $S6K1^{-/-}$  mice were more insulin sensitive, as indicated by faster glucose clearance in the

insulin tolerance test (Fig. 3a). In contrast, both genotypes displayed increased hyperglycaemia on a HFD, although the effect was more pronounced in wild-type mice (Fig. 3b). However, despite the increases in glucose and in FFAs,  $S6K1^{-/-}$  mice remain as insulin sensitive on a HFD as on a NCD (Fig. 3b versus a). In the case of wild-type mice, insulin resistance on a HFD (Fig. 3b) can be explained by persistently elevated insulin levels (Table 1), inducing insulin receptor desensitization, as measured by reduced receptor auto-phosphorylation in response to insulin (Fig. 3c). In  $S6K1^{-/-}$  mice insulin levels fail to rise on a HFD (Table 1), consistent with higher insulin receptor auto-phosphorylation (Fig. 3c). However, insulin receptors still desensitize in  $S6K1^{-/-}$  mice on a HFD (Fig. 3c), probably due to the large increase in FFA levels (Table 1). That insulin receptors desensitize in  $S6K1^{-/-}$  mice but remain insulin sensitive suggests that absence of S6K1 facilitates insulin signalling downstream of the insulin receptor. To test this we monitored protein kinase B (PKB) phosphorylation, as insulin sensitivity is tightly linked to the phosphatidylinositol-3-OH kinase (PI(3)K)–PKB signalling pathway<sup>14</sup>. As with insulin receptor auto-phosphorylation in wild-type mice on a HFD, insulin-induced PKB phosphorylation was suppressed in fat, liver and muscle compared with wild-type mice maintained on a NCD (Fig. 3d). However, there was no significant effect on PKB phosphorylation in  $S6K1^{-/-}$  mice,



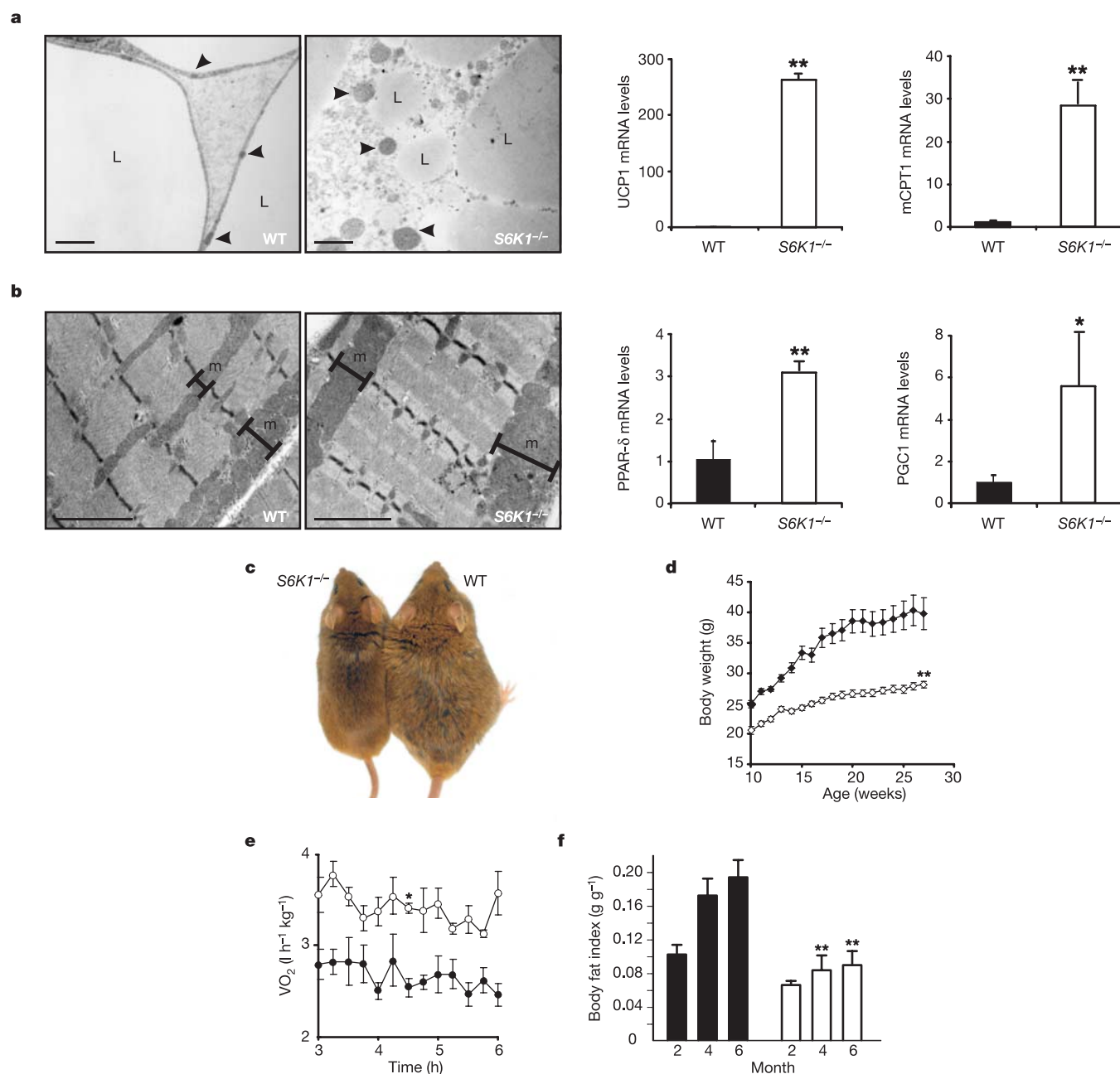
**Figure 1** Reduced adiposity in  $S6K1^{-/-}$  mice. **a**, Growth curves of wild-type and  $S6K1^{-/-}$  mice on a NCD. Wild type,  $n = 11$ ;  $S6K1^{-/-}$ ,  $n = 13$ . **b**, Weights of epididymal, inguinal and retroperitoneal adipose tissue (EWAT, IWAT and RWAT, respectively) and liver normalized by body weight ( $n = 6$  each genotype). WT, wild type. **c**, Scanning electron microscopic (top) and histological analysis of epididymal WAT (bottom). Arrowheads indicate multi-locular adipocytes. Magnification for scanning microscopy and for histology are  $\times 500$  and  $\times 200$ , respectively. Mice were 6-month-

old males in **b** and **c**. **d**, Food intake per mouse measured over 15 days ( $n = 12$ ;  $P = 0.8$ ) or normalized by body weight ( $n = 12$ ). **e**, Enhanced lipolysis in  $S6K1^{-/-}$  adipocytes. **f**, Oxygen consumption (VO<sub>2</sub>) in wild-type and  $S6K1^{-/-}$  mice fed a NCD ( $n = 6$  each genotype). All values are given as mean  $\pm$  s.e.m. Asterisk,  $P < 0.01$ ; double asterisk,  $P < 0.001$ . Filled symbols/columns indicate wild type; open symbols/columns,  $S6K1^{-/-}$  mice.

regardless of diet or tissue (Fig. 3d). This suggested that S6K1 elicited a selective inhibitory effect on PKB activation at a point downstream of the insulin receptor, consistent with little effect on mitogen activated protein kinase and S6K1 activation in wild-type animals on a HFD (see below and data not shown).

Recently, we demonstrated in *Drosophila* that dS6K negatively regulates dPKB activity in a cell-autonomous manner<sup>15</sup>. To test whether this was the mechanism responsible for the observed

responses, S6K1 levels were reduced with small interfering RNAs. The results show that lowering S6K1 levels potentiates insulin-induced PKB phosphorylation, with no effect on insulin receptor auto-phosphorylation (Fig. 3e), suggesting that the target of inhibition resided downstream of the insulin receptor. Indeed, reduction of S6K1 levels is paralleled by a decrease in phosphorylation of S307 and S636/S639 of insulin receptor substrate 1 (IRS1) (Fig. 3e), sites shown to inhibit PI(3)K binding to IRS1 (ref. 4) and



**Figure 2** Increased mitochondria and resistance to diet-induced obesity. **a**, The left panels show transmission electron microscopic analysis of epididymal fat. Arrowheads indicate mitochondria and lipid droplets (L) ( $\times 5,000$  magnification). The right panels show UCP1 and mCPT1 mRNA (relative values) measured by quantitative RT-PCR (see Methods). **b**, The left panels show transmission electron microscopic analysis of plantaris muscle ( $\times 10,000$  magnification). m, mitochondria. The right panels show PPAR- $\delta$  and PGC1 mRNA levels (relative values) measured by quantitative RT-PCR. In **a**, **b**,  $n = 4-7$ ; asterisk,  $P < 0.01$ ; double asterisk,  $P < 0.001$ . **c**, Representative wild-type and

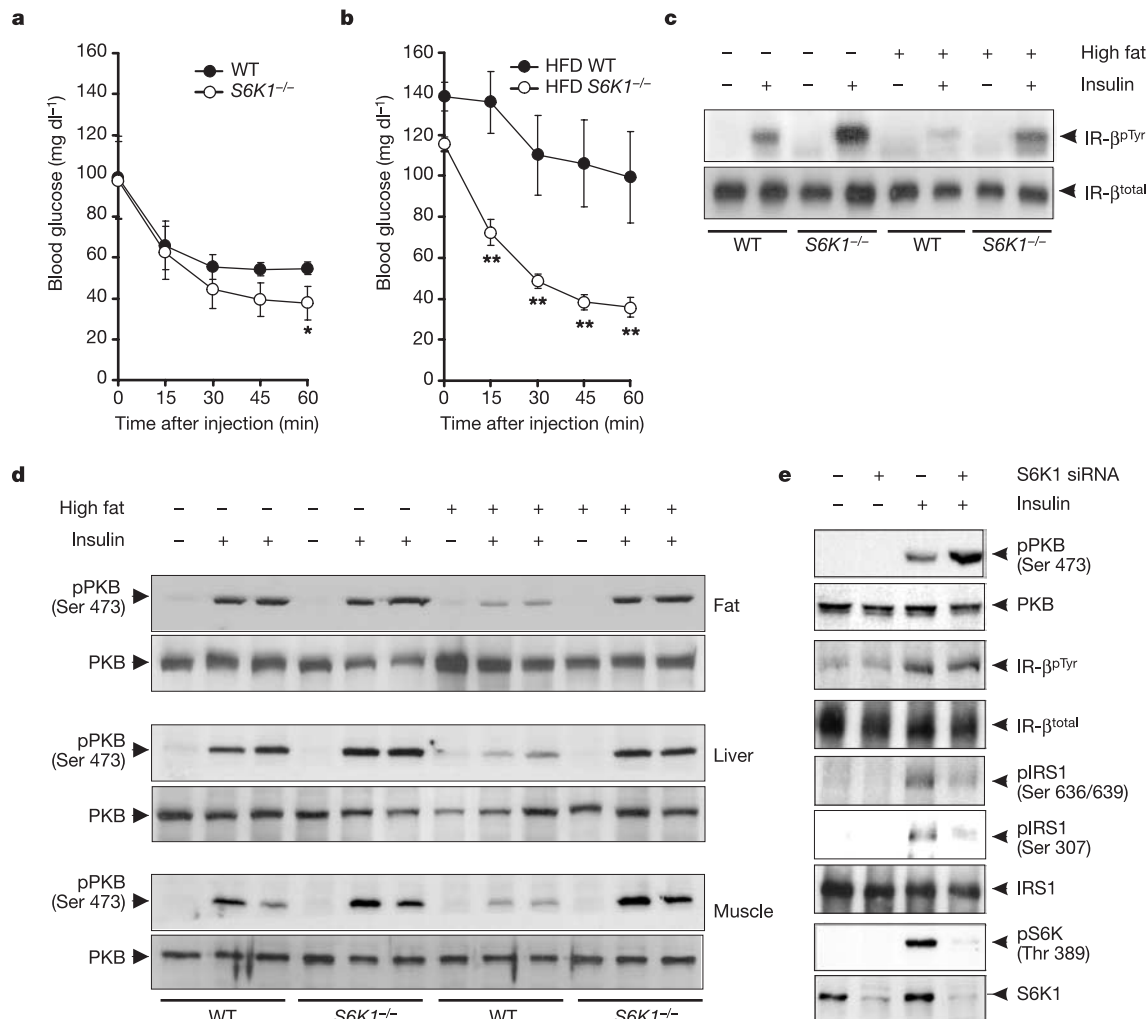
*S6K1*<sup>-/-</sup> mice after 6 months feeding on a HFD. **d**, Growth curves of wild-type and *S6K1*<sup>-/-</sup> mice maintained on a HFD. WT,  $n = 11$ ; *S6K1*<sup>-/-</sup>,  $n = 15$ ;  $P < 0.001$ . **e**, Oxygen consumption ( $VO_2$ ) in wild-type and *S6K1*<sup>-/-</sup> mice maintained on a HFD for 6 months ( $n = 6$  each genotype). **f**, Body fat index determined by MRI analysis of wild-type ( $n = 10$ ) and *S6K1*<sup>-/-</sup> mice ( $n = 12$ ) maintained on a HFD for 4 months. Filled symbols/columns indicate wild type; open symbols/columns, *S6K1*<sup>-/-</sup> mice. Values are mean  $\pm$  s.e.m. Asterisk,  $P < 0.01$ ; double asterisk,  $P < 0.001$  in **d-f**.

to be involved in insulin resistance in skeletal muscle cells from patients with type 2 diabetes<sup>5</sup>. Thus, removal of S6K1 can facilitate insulin signalling in a cell-autonomous manner.

Hyperglycaemia, hyperaminoacidaemia and hyperlipidaemia are associated with obesity and insulin resistance<sup>16</sup>; however, the role of increased nutrients in insulin action is not well understood<sup>17–19</sup>. As S6K1 is activated by nutrients<sup>20–22</sup> and acts negatively on PI(3)K signalling, this raised the possibility that on a HFD S6K1 is involved in inducing insulin resistance. This hypothesis is supported by the reversal of amino acid inhibition of insulin-induced PI(3)K signalling by rapamycin, which inhibits mTOR<sup>23</sup>, an immediate upstream S6K1 kinase<sup>2</sup>. Consistent with this hypothesis, phosphorylation of S6K1 T389, S6 S240/S244, IRS1 S307 and IRS1 S636/S639 is highly elevated in wild-type mice maintained on a HFD compared with wild-type mice maintained on a NCD (Fig. 4a, b). Furthermore, the increase in IRS1 S307 and S636/S639 phosphorylation is absent in *S6K1*<sup>−/−</sup> mice on a HFD (Fig. 4b). Under these conditions there were no apparent alterations in IRS1 levels (Fig. 4b). These findings suggest that nutrient-induced S6K1 activation acts to suppress insulin signalling through modulating IRS1 S307 and S636/S639 phosphorylation. To test this further, two genetic models of obesity were examined: *K/K A<sup>y</sup>* and *ob/ob* mice<sup>24,25</sup>. The results show that

the *K/K A<sup>y</sup>* and *ob/ob* mice maintained on a NCD have elevated S6K1 T389, S6 S240/S244, IRS1 S307 and IRS1 S636/S639 phosphorylation as compared with wild-type mice on a NCD (Fig. 4c). Moreover, in contrast with PKB S473 phosphorylation in adipose and muscle (Fig. 3d), insulin stimulates S6K1 T389 phosphorylation to even higher levels in wild-type animals on a HFD compared with a NCD (Fig. 4d), potentially further suppressing PI(3)K signalling. Thus, either nutritionally or genetically driven obesity leads to the upregulation of S6K1, which may in turn act to suppress PI(3)K signalling, contributing to insulin resistance.

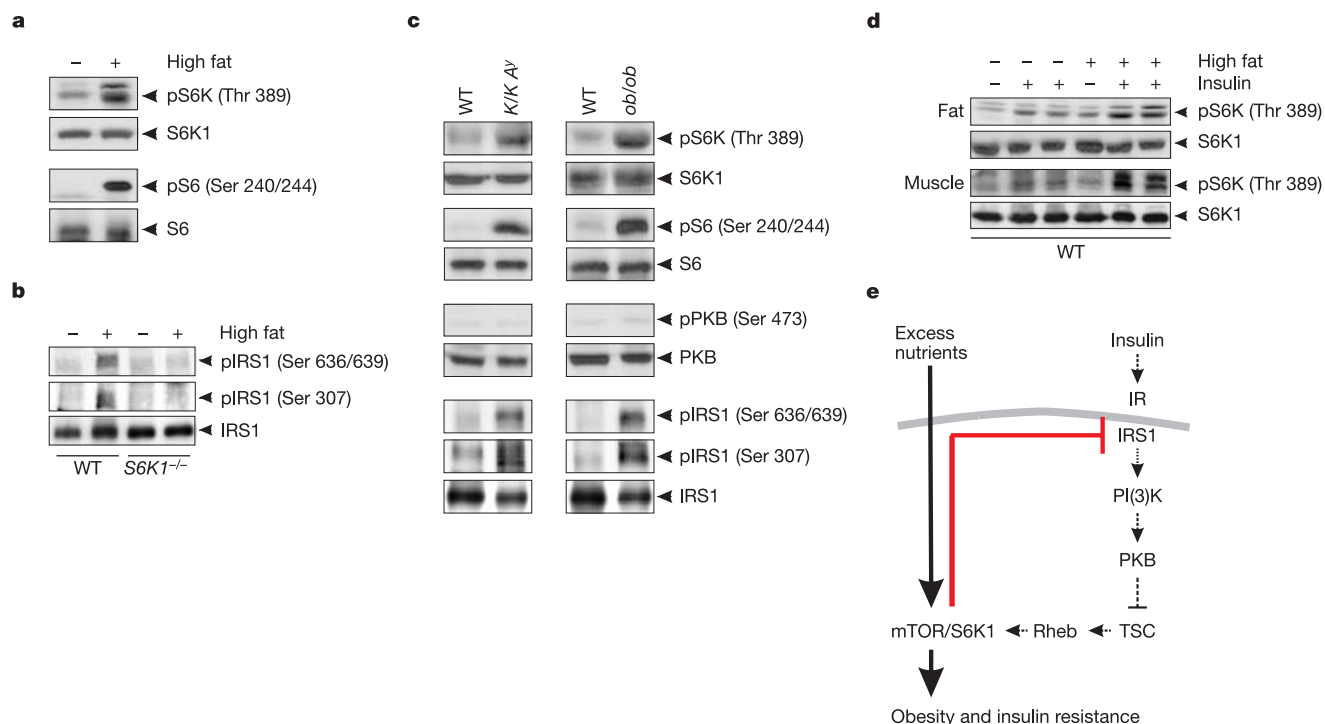
The results presented here indicate that *S6K1*<sup>−/−</sup> mice are protected against obesity and insulin resistance due to the upregulation of the oxidative phosphorylation pathway and increased insulin sensitivity. Enhanced oxidative metabolism is consistent with the increase in mitochondria number, as well as the induction of genes that control the oxidative phosphorylation pathway. That *S6K1*<sup>−/−</sup> mice remain insulin sensitive despite high circulating FFAs may be explained by the strong protection against metabolic syndrome by overexpression or activation of PPAR- $\delta$ <sup>6–8</sup>. Despite the increase in the oxidative phosphorylation pathway and reduced insulin levels, circulating glucose and FFA levels still rise in *S6K1*<sup>−/−</sup> mice; however, these animals remain sensitive to insulin, and PI(3)K



**Figure 3** Enhanced insulin sensitivity and insulin signalling in the absence of S6K1. **a**, **b**, Insulin tolerance test in 3-h-fasted mice on a NCD (**a**) or HFD (**b**) at 6 months of age. Asterisk,  $P < 0.05$ ; double asterisk,  $P < 0.01$ . Values are mean  $\pm$  s.e.m. **c**, Insulin receptor tyrosine phosphorylation in liver of wild-type and *S6K1*<sup>−/−</sup> mice before (−) or

after (+) insulin stimulation on a NCD or HFD. **d**, Level of phosphorylation of PKB Ser 473 in fat, liver and muscle ( $n = 3–8$  for each genotype in **c**, **d**). **e**, Lowering S6K1 expression by siRNA potentiates insulin-induced PKB phosphorylation in HeLa cells.





**Figure 4** S6K1 activation in obesity. **a**, S6K1 Thr 389 and S6 Ser 240/244 phosphorylation in fat of wild-type mice maintained on a NCD (–) or HFD (+) for 4 months. **b**, IRS1 Ser 307 and Ser 636/639 phosphorylation in fat of wild-type or *S6K1*<sup>–/–</sup> mice maintained on a NCD (–) or HFD (+). **c**, S6K1 Thr 389, S6 Ser 240/244, IRS1 Ser 307 and IRS1 Ser 636/639 phosphorylation in fat of wild-type, *K/K A<sup>Y</sup>* or *ob/ob* mice maintained on a NCD. **d**, S6K1 Thr 389 phosphorylation in fat or muscle of wild-type

mice maintained on a NCD or HFD for 4 months before (–) or after (+) injection of insulin (0.75 U kg<sup>–1</sup> of body weight). All mice of each genotype (*n* = 3–8) in **a–d** were fasted for 6 h and were 6 months old. **e**, Model of inhibition of IRS1 signalling through activation of S6K1.

signalling is unaffected. This observation may be explained by the loss of S6K1, whose activation by either nutrients or insulin leads to increased IRS1 serine phosphorylation (Fig. 4e). In the case of insulin, this is mediated by a negative feedback loop triggered by PKB phosphorylation of the TSC1/2 tumour suppressor complex, leading to Rheb activation and stimulation of S6K1 (ref. 26). Thus, in a homeostatic setting, as nutrients and amino acids are consumed, mTOR/S6K1 activity would decrease, restoring PI(3)K signalling, whereas the incessant supply of nutrients associated with the obese state would lead to constitutive activation of mTOR/S6K1 and desensitization of insulin signalling (Fig. 4e). Taken together the results suggest that S6K1 may have a central function along with other signalling components in development of obesity and insulin resistance, and may be an important drug target in the treatment of patients suffering from these pathological disorders. □

## Methods

### Mice

*S6K1*<sup>–/–</sup> mice were generated as previously described<sup>3</sup>. Male C57BL/6J, *K/K A<sup>Y</sup>* and *ob/ob* (C57BL/6J background) mice were obtained from E. Janvier, CLEA Japan Inc., and The Jackson Laboratory, respectively.

### Metabolic studies

At 10 weeks of age male mice were placed on either a NCD (diet number 3807, KLIBA-NAFAG) or HFD *ad libitum* (diet D12492, Researchdiets) and monitored for 24 weeks. Body weight was recorded weekly and food intake was measured every second day for 15 consecutive days. Insulin tolerance tests, oxygen consumption, RER measurements and quantification of blood metabolites were performed as previously described<sup>27</sup>.

### Histology and morphometric analysis of tissues

Adipose tissue was analysed by haematoxylin and eosin staining as previously described<sup>27</sup>. Morphometric analysis of epididymal WAT from 500 or more cells from three different

animals per genotype was performed with ImageJ software (NIH). Adipose and plantaris muscle tissue were prepared as described for scanning and transmission electron microscopy<sup>27</sup>.

### Magnetic resonance imaging analysis

MRI experiments were carried out on a Biospec 47/30 spectrometer (Bruker Medical) at 4.7 T equipped with a self-shielded 12-cm bore gradient system<sup>28</sup>. Animals were anaesthetized with 1.5% isoflurane (Abbott). Adipose tissue was measured with an optimized turbo-RARE2 imaging sequence. Acquisition parameters were: repetition delay (TR) = 250 ms; echo delay (TE) = 8.6 ms; RARE factor = 32 (effective echo time 73.1 ms); number of averages = 8; slice orientation transverse, image matrix = 128 × 128 pixels; field-of-view = 3.5 × 3.5 cm; slice thickness = 1.2 mm (contiguous). Fat pad volumes were assessed with an in-house software algorithm based on IDL software package (Research Systems Inc.). Body fat indices were calculated by dividing adipose tissue weight by body weight.

### Lipolysis in isolated adipocytes

Primary adipocytes were prepared from epididymal fat pads as described previously<sup>29</sup>. Cells were incubated for 30 min at 37 °C with or without norepinephrine (Sigma-Aldrich SARL) at the indicated concentrations.

### Real-time quantitative RT-PCR

Total RNA was extracted from frozen tissue samples or cells using the RNeasy kit (Qiagen). Complementary DNA was synthesized from total RNA with the SuperScript First-Strand Synthesis System (Invitrogen) and random hexamer primers. The real-time polymerase chain reaction (PCR) measurement of individual cDNAs was performed using SYBR green dye to measure duplex DNA formation with the Roche Lightcycler system and normalized to the expression of either  $\beta$ -actin or 18S ribosomal RNA. The primers and probes used in the real time RT-PCR were the following: UCP1 sense 5'-GGCCCTTGTAAACAACAAATAC-3', antisense 5'-GGCAACAAGAGCTGACAGTAAAT-3'; UCP3 sense 5'-ACTCCAGCGTCGCCATCAGGATTCT-3', antisense 5'-TAAACAGGTGAGACTCCAGCAACTT-3'; mCPT1 sense 5'-TTGCCCTACAGCTCTGGCAITTTCC-3', antisense 5'-GCACCCAGATGATTTGGGATACTGT-3'; mPPAR- $\delta$  sense 5'-CTCTTCATCGCGGCCATCATTCT-3', antisense 5'-TCTGCCATCTTCTGCAGCAGCTT-3'; PGC-1 sense 5'-AAGTGTGGAACCTCTCTGGAAGT-3', antisense 5'-GGGTTATCTTGGTTGGCTTATG-3'.

# Measurement of insulin receptor and IRS1 phosphorylation *in vivo*

After a 6-h fast mice were injected intravenously with 0.75 U kg<sup>-1</sup> insulin (Eli Lilly) or equal volume of vehicle. All indicted tissues were collected in liquid nitrogen 5 min after injection. Protein extracts from tissue samples were analysed as described<sup>30</sup>. Antibodies were from Santa Cruz (anti-insulin receptor  $\beta$  and anti-S6K1 antibodies), Upstate Biotechnology (anti-phosphotyrosine and anti-IRS1 antibodies) and Cell Signaling (anti-PKB, anti-phospho-PKB Ser 473, anti-phospho-IRS1 Ser 636/639, anti-phospho-S6K Thr 389 and anti-phospho-S6 240/244 antibodies). Antibodies to S6 and phospho-IRS1 Ser 307 were from J. Mestan and Y. Le Marchand-Brustel, respectively.

## RNA interference

RNA interference (RNAi) duplexes corresponding to human S6K1 (5'-AAGGGGCTATGGAAAGGTTT-3') were purified, annealed and transfected into HeLa cells using oligofectamine (Invitrogen). After 60 h cells were deprived of serum overnight and either lysed directly or stimulated with 200 nM insulin for 30 min. The effect of RNAi on S6K1 expression and PKB phosphorylation was measured by western blot analysis. Cell lysates were incubated for 4 h with anti-IRS1 or anti-insulin receptor  $\beta$  antibody pre-absorbed on protein A Sepharose at 4 °C, and analysed by western blot analyses after gel electrophoresis.

## Statistical analysis

Data are presented as mean  $\pm$  s.e.m. The main and interactive effects were analysed by analysis of variance (ANOVA) factorial, repeated measurements or by one-way ANOVA followed by Bonferroni *t*-test (MRI analysis). Differences between individual group means were analysed by Fisher's PLSD test. Analyses were performed using Statview Software (Brainpower).

Received 19 April; accepted 21 July 2004; doi:10.1038/nature02866.

Published online 11 August 2004.

1. Saltiel, A. R. & Kahn, C. R. Insulin signalling and the regulation of glucose and lipid metabolism. *Nature* **414**, 799–806 (2001).
2. Fingar, D. C., Salama, S., Tsou, C., Harlow, E. & Blenis, J. Mammalian cell size is controlled by mTOR and its downstream targets S6K1 and 4EBP1/eIF4E. *Genes Dev.* **16**, 1472–1487 (2002).
3. Pende, M. *et al.* Hypoinsulinaemia, glucose intolerance and diminished  $\beta$ -cell size in S6K1-deficient mice. *Nature* **408**, 994–997 (2000).
4. Zick, Y. Insulin resistance: a phosphorylation-based uncoupling of insulin signaling. *Trends Cell Biol.* **11**, 437–441 (2001).
5. Bouzakri, K. *et al.* Reduced activation of phosphatidylinositol-3 kinase and increased serine 636 phosphorylation of insulin receptor substrate-1 in primary culture of skeletal muscle cells from patients with type 2 diabetes. *Diabetes* **52**, 1319–1325 (2003).
6. Tanaka, T. *et al.* Activation of peroxisome proliferator-activated receptor delta induces fatty acid  $\beta$ -oxidation in skeletal muscle and attenuates metabolic syndrome. *Proc. Natl Acad. Sci. USA* **100**, 15924–15929 (2003).
7. Luquet, S. *et al.* Peroxisome proliferator-activated receptor delta controls muscle development and oxidative capability. *FASEB J.* **17**, 2299–2301 (2003).
8. Dressel, U. *et al.* The peroxisome proliferator-activated receptor beta/delta agonist, GW501516, regulates the expression of genes involved in lipid catabolism and energy uncoupling in skeletal muscle cells. *Mol. Endocrinol.* **17**, 2477–2493 (2003).
9. Patti, M. E. *et al.* Coordinated reduction of genes of oxidative metabolism in humans with insulin resistance and diabetes: Potential role of PGC1 and NRF1. *Proc. Natl Acad. Sci. USA* **100**, 8466–8471 (2003).
10. Mootha, V. K. *et al.* PGC-1 $\alpha$ -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genet.* **34**, 267–273 (2003).
11. Boden, G. Role of fatty acids in the pathogenesis of insulin resistance and NIDDM. *Diabetes* **46**, 3–10 (1997).
12. Kahn, B. B. Type 2 diabetes: when insulin secretion fails to compensate for insulin resistance. *Cell* **92**, 593–596 (1998).
13. Le Marchand-Brustel, Y. *et al.* Fatty acid-induced insulin resistance: role of insulin receptor substrate 1 serine phosphorylation in the retroregulation of insulin signalling. *Biochem. Soc. Trans.* **31**, 1152–1156 (2003).
14. Jiang, G. & Zhang, B. B. Pi 3-kinase and its up- and down-stream modulators as potential targets for the treatment of type II diabetes. *Front. Biosci.* **7**, d903–d907 (2002).
15. Radimerski, T., Montagne, J., Hemmings-Mieszcak, M. & Thomas, G. Lethality of *Drosophila* lacking TSC tumor suppressor function rescued by reducing dS6K signaling. *Genes Dev.* **16**, 2627–2632 (2002).
16. Kahn, B. B. & Flier, J. S. Obesity and insulin resistance. *J. Clin. Invest.* **106**, 473–481 (2000).
17. Proietto, J., Filippis, A., Nakhl, C. & Clark, S. Nutrient-induced insulin resistance. *Mol. Cell. Endocrinol.* **151**, 143–149 (1999).
18. Greiwe, J. S., Kwon, G., McDaniel, M. L. & Semenkovich, C. F. Leucine and insulin activate p70 S6 kinase through different pathways in human skeletal muscle. *Am. J. Physiol. Endocrinol. Metab.* **281**, E466–E471 (2001).
19. Patti, M. E. Nutrient modulation of cellular insulin action. *Ann. NY Acad. Sci.* **892**, 187–203 (1999).
20. Hara, K. *et al.* Amino acid sufficiency and mTOR regulate p70 S6 kinase and eIF-4E BP1 through a common effector mechanism. *J. Biol. Chem.* **273**, 14484–14494 (1998).
21. Patti, M. E., Brambilla, E., Luzzi, L., Landaker, E. J. & Kahn, C. R. Bidirectional modulation of insulin action by amino acids. *J. Clin. Invest.* **101**, 1519–1529 (1998).
22. Dennis, P. B. *et al.* Mammalian TOR: a homeostatic ATP sensor. *Science* **294**, 1102–1105 (2001).
23. Tremblay, F. & Marette, A. Amino acid and insulin signaling via the mTOR/p70 S6 kinase pathway: A negative feedback mechanism leading to insulin resistance in skeletal muscle cells. *J. Biol. Chem.* **276**, 38052–38060 (2001).
24. Iwatsuka, H., Shino, A. & Suzuoki, Z. General survey of diabetic features of yellow KK mice. *Endocrinol. Jpn* **17**, 23–35 (1970).
25. Zhang, Y. *et al.* Positional cloning of the mouse obese gene and its human homologue. *Nature* **372**, 425–432 (1994).

26. Manning, B. D. & Cantley, L. C. Rheb fills a GAP between TSC and TOR. *Trends Biochem. Sci.* **28**, 573–576 (2003).
27. Picard, F. *et al.* SRC-1 and TIF2 control energy balance between white and brown adipose tissues. *Cell* **111**, 931–941 (2002).
28. Doty, F. D., Entzinger, G. Jr, Hauck, C. D. & Staab, J. P. Practical aspects of birdcage coils. *J. Magn. Reson.* **138**, 144–154 (1999).
29. Marette, A., Tulp, O. L. & Bukowiecki, L. J. Mechanism linking insulin resistance to defective thermogenesis in brown adipose tissue of obese diabetic SHR/N-cp rats. *Int. J. Obes.* **15**, 823–831 (1991).
30. Hirosumi, J. *et al.* A central role for JNK in obesity and insulin resistance. *Nature* **420**, 333–336 (2002).

Supplementary Information accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank T. Opgenorth and C. Rondinone for sharing their results before publication; G. S. Hotamisligil, S. Y. Kim and D. J. Withers for their critical reading of the manuscript; and S. Cinti, P. B. Dennis, A. Dullloo, L. Fajas, A. Greenberg, B. M. Spiegelman, G. Solinas, J. Tanti and M. Wymann for discussions. We are also grateful to M.-F. Champy, W. Theilkaes, N. Messaddeq, I. Obergfoell and J. F. Spetz for the blood analysis, studies with MRI, technical assistance with electron microscopy, for photography and for assistance in the animal experiments, respectively. Work in the laboratory of J.A. is supported by grants from CNRS, INSERM, ULP, Hôpital Universitaire de Strasbourg, NIH, EMBO and the European community, and the laboratory of S.C.K. and G.T. is supported by the Novartis Institutes for Biomedical Research and a grant from the Swiss Cancer League.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to G.T. ([gthomas@fmi.ch](mailto:gthomas@fmi.ch)).

## Cytoplasmic PML function in TGF- $\beta$ signalling

Hui-Kuan Lin, Stephan Bergmann & Pier Paolo Pandolfi

Cancer Biology and Genetics Program, Department of Pathology, Memorial Sloan-Kettering Cancer Center, Sloan-Kettering Institute, 1275 York Avenue, New York, New York, 10021, USA

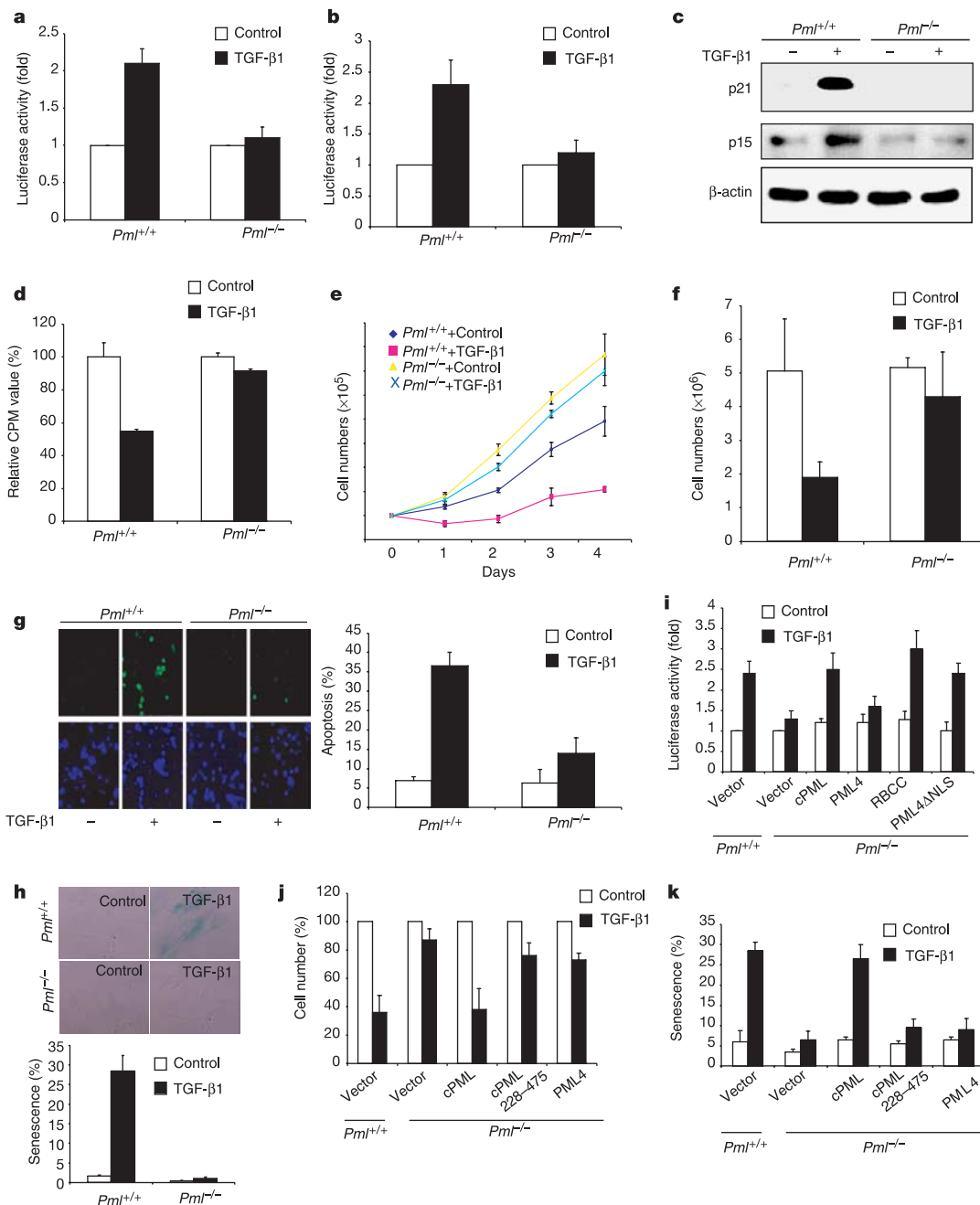
Transforming growth factor  $\beta$  (TGF- $\beta$ ) is a pluripotent cytokine that controls key tumour suppressive functions<sup>1–3</sup>, but cancer cells are often unresponsive to it<sup>1,4</sup>. The promyelocytic leukaemia (PML) tumour suppressor of acute promyelocytic leukaemia (APL) accumulates in the PML nuclear body, but cytoplasmic PML isoforms of unknown function have also been described<sup>5,6</sup>. Here we show that cytoplasmic *Pml* is an essential modulator of TGF- $\beta$  signalling. *Pml*-null primary cells are resistant to TGF- $\beta$ -dependent growth arrest, induction of cellular senescence and apoptosis. These cells also have impaired phosphorylation and nuclear translocation of the TGF- $\beta$  signalling proteins Smad2 and Smad3, as well as impaired induction of TGF- $\beta$  target genes. Expression of cytoplasmic *Pml* is induced by TGF- $\beta$ . Furthermore, cytoplasmic PML physically interacts with Smad2/3 and SARA (Smad anchor for receptor activation) and is required for association of Smad2/3 with SARA and for the accumulation of SARA and TGF- $\beta$  receptor in the early endosome. The PML-RAR $\alpha$  oncoprotein of APL can antagonize cytoplasmic PML function and APL cells have defects in TGF- $\beta$  signalling similar to those observed in *Pml*-null cells. Our findings identify cytoplasmic PML as a critical TGF- $\beta$  regulator, and further implicate deregulated TGF- $\beta$  signalling in cancer pathogenesis.

APL is almost invariably associated with chromosomal translocations involving the *PML* tumour suppressor and *RAR $\alpha$*  genes, resulting in the generation of a PML-RAR $\alpha$  leukaemogenic fusion protein that can function as a dominant-negative PML and RAR $\alpha$  mutant<sup>7–9</sup>. PML-RAR $\alpha$  physically interacts with nuclear PML isoforms causing their delocalization from the PML nuclear body (PML-NB) into aberrant microspeckled nuclear structures<sup>7,8</sup>. Until

now, the tumour suppressive role of PML has been attributed to its ability to regulate the transcriptional function of tumour suppressors, such as p53 and Rb, within the nucleus<sup>8</sup>. However, several cytoplasmic PML (cPML) isoforms, of unknown function, have been described. Here, we show that cPML has an essential function

in the modulation of TGF- $\beta$  signalling in the cytosolic fraction that is profoundly impaired in the APL blasts.

We first examined whether PML functions in TGF- $\beta$ -induced gene expression. In *Pml*<sup>-/-</sup> mouse embryonic fibroblasts (MEFs), TGF- $\beta$ 1 activation of the PF-1-luciferase reporter (PF-1-luc), a



**Figure 1** Pml is essential for TGF- $\beta$ 1 transcriptional and biological functions. Wild-type or *Pml*<sup>-/-</sup> MEFs were co-transfected with Smad3, PRL-SV40-luc and a PF-1-luc reporter (**a**) or a 3PT-luc reporter (**b**). Cells were either left untreated or were treated with TGF- $\beta$ 1 (100 pM) for 30 h and then harvested for a luciferase activity assay. **c**, MEFs were treated with TGF- $\beta$ 1 for 24 h and harvested for western blot analysis. **d**, MEFs were pretreated with or without TGF- $\beta$ 1 12 h prior to labelling with <sup>3</sup>H-thymidine for 12 h, and cells were harvested for thymidine incorporation assays. The proportion of cells incubated without TGF- $\beta$ 1 was defined as 100%. **e**, MEFs were either left untreated or were treated with TGF- $\beta$ 1, harvested, and stained with trypan blue at the various times, as indicated. TGF- $\beta$ 1 was re-added on alternate days. **f**, B cells isolated from the wild-type or *Pml*<sup>-/-</sup> mouse were either left untreated or were treated with TGF- $\beta$ 1 for 30 h, and cells were

stained with trypan blue. **g**, B cells were either left untreated or were treated with TGF- $\beta$ 1 for 30 h, and apoptosis was determined by TUNEL assay. The top and bottom panels are TUNEL staining and DAPI staining, respectively. **h**, MEFs were either left untreated or were treated with TGF- $\beta$ 1 for 6 days, harvested and stained with  $\beta$ -gal. **i**, Same as in **a**, except cells were transfected with plasmids as indicated. **j**, MEFs infected with the indicated retroviral expression vectors, were either left untreated or were treated with TGF- $\beta$ 1 for 4 days. Cells were stained with trypan blue and counted for viability as in **e**. **k**, MEFs infected with the indicated retroviral expression vectors, were either left untreated or were treated with TGF- $\beta$ 1 for 6 days, harvested and stained with  $\beta$ -gal. The results are presented as mean values  $\pm$  standard deviation (s.d.) from a representative experiment performed in triplicate.



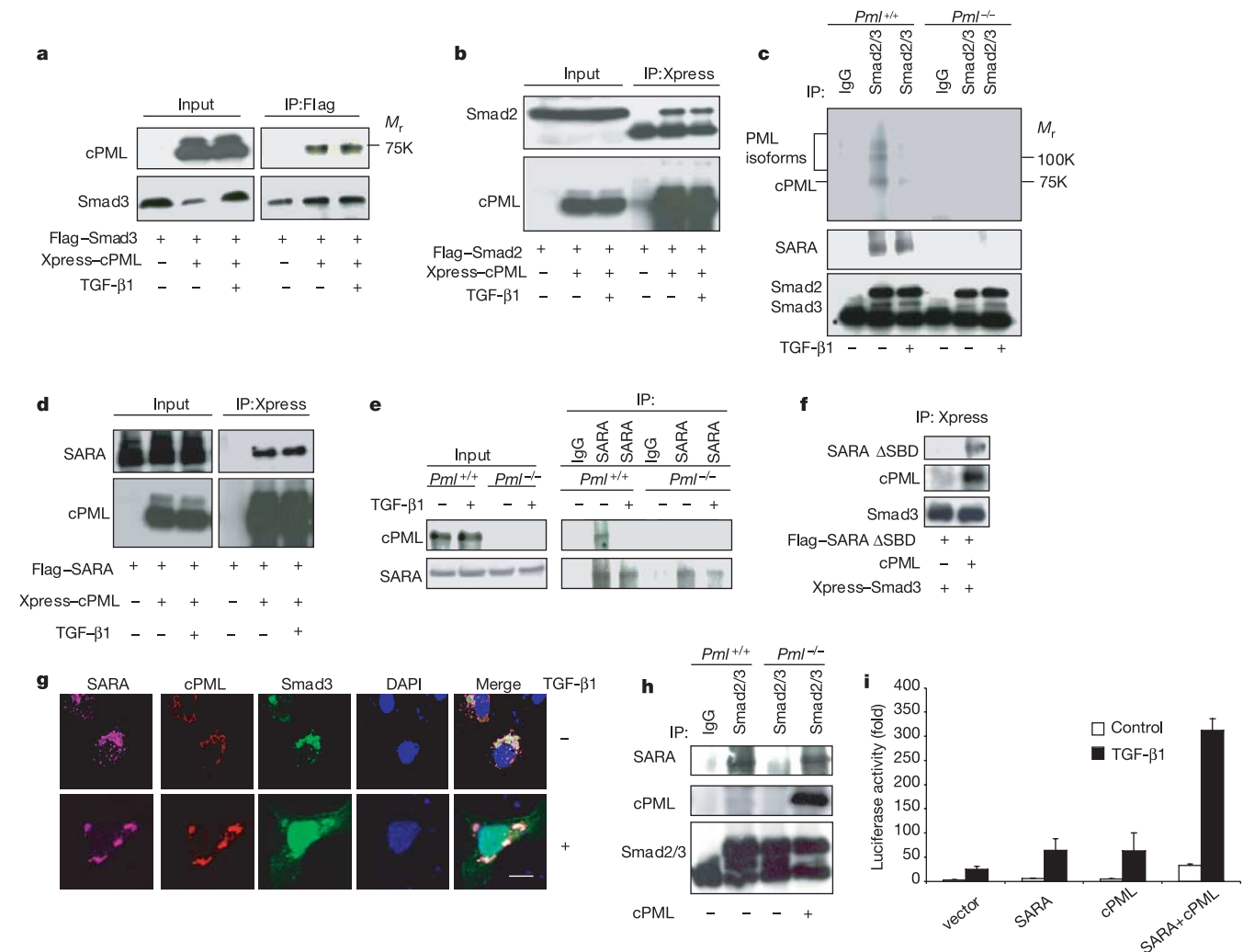
promoter of plasminogen activator inhibitor-1 (PAI-1) and 3TP-luc was markedly impaired (Fig. 1a,b). To further confirm the defect in TGF- $\beta$ -induced gene expression, we tested the effects of TGF- $\beta$  on endogenous target gene expression. Two cyclin inhibitors, p21 and p15, that are known to mediate TGF- $\beta$ -induced growth arrest<sup>1,2</sup>, were readily upregulated in *Pml*<sup>+/+</sup>, but not in *Pml*<sup>-/-</sup>, MEFs (Fig. 1c).

As loss of Pml impairs TGF- $\beta$  target gene expression, we tested whether TGF- $\beta$ -mediated biological functions, such as cell growth inhibition and apoptosis, may be impaired after *Pml* inactivation. Indeed, TGF- $\beta$ 1 growth inhibitory function was markedly reduced both in MEFs and B cells isolated from *Pml*<sup>-/-</sup> mice (Fig. 1d-f). Similarly, TGF- $\beta$ 1-dependent apoptosis in *Pml*-null B cells was also profoundly impaired (Fig. 1g).

PML has been implicated in the induction of cellular senescence after oncogenic transformation<sup>10,11</sup>. TGF- $\beta$  has also been shown to induce senescence in cancer cells<sup>12</sup>. We therefore examined whether TGF- $\beta$  induces senescence in MEFs, and whether PML modulates this function. Notably, TGF- $\beta$ 1 functioned as a powerful inducer of

cellular senescence in wild-type MEFs, whereas the effect was completely abrogated in *Pml*<sup>-/-</sup> MEFs (Fig. 1h).

We next examined whether add-back of PML isoforms in *Pml*<sup>-/-</sup> MEFs rescues TGF- $\beta$ -induced transcriptional activity and biological functions. Restoration of PML4 (a nuclear PML isoform) expression in *Pml*<sup>-/-</sup> MEFs only partially rescued TGF- $\beta$ -induced gene expression (Fig. 1i). Notably, add-back of cPML (PML3 3-7; ref. 6) fully rescued the TGF- $\beta$ 1-mediated transcriptional activity (Fig. 1i). We next tested whether the cytoplasmic localization of a nuclear PML isoform could also rescue TGF- $\beta$ 1 unresponsiveness. For this, we used a PML4 mutant devoid of a nuclear localization signal (PML4  $\Delta$ NLS) and PML4 RBCC (ring/B-box/coiled coil region) that entirely lacks the carboxyl terminus. Both mutants displayed a cytoplasmic localization (data not shown) and restored TGF- $\beta$ -induced transcriptional activity (Fig. 1i). We next tested whether add-back of cPML would rescue TGF- $\beta$ 1-induced cell growth inhibition and senescence. cPML readily rescued these functional losses, whereas neither PML4 nor a cPML mutant



**Figure 2** cPML interacts physically and functionally with Smad2/3 and SARA. **a, b**, 293T cells were transfected with the indicated plasmids for 48 h, treated with or without TGF- $\beta$ 1 for 1 h, and harvested for co-immunoprecipitation (Co-IP) experiments. **c**, Total-cell extracts prepared from MEFs treated with or without TGF- $\beta$ 1 for 1 h were subjected to immunoprecipitation with IgG or Smad2/3 antibodies, followed by western blot analysis. **d**, 293T cells were transfected with plasmids, as indicated, for 48 h, treated with or without TGF- $\beta$ 1 for 1 h and harvested for Co-IP experiments. **e**, Total-cell extracts prepared from MEFs treated with or without TGF- $\beta$ 1 for 1 h were subjected to immunoprecipitation with SARA antibody, followed by western blot analysis. **f**, 293T cells

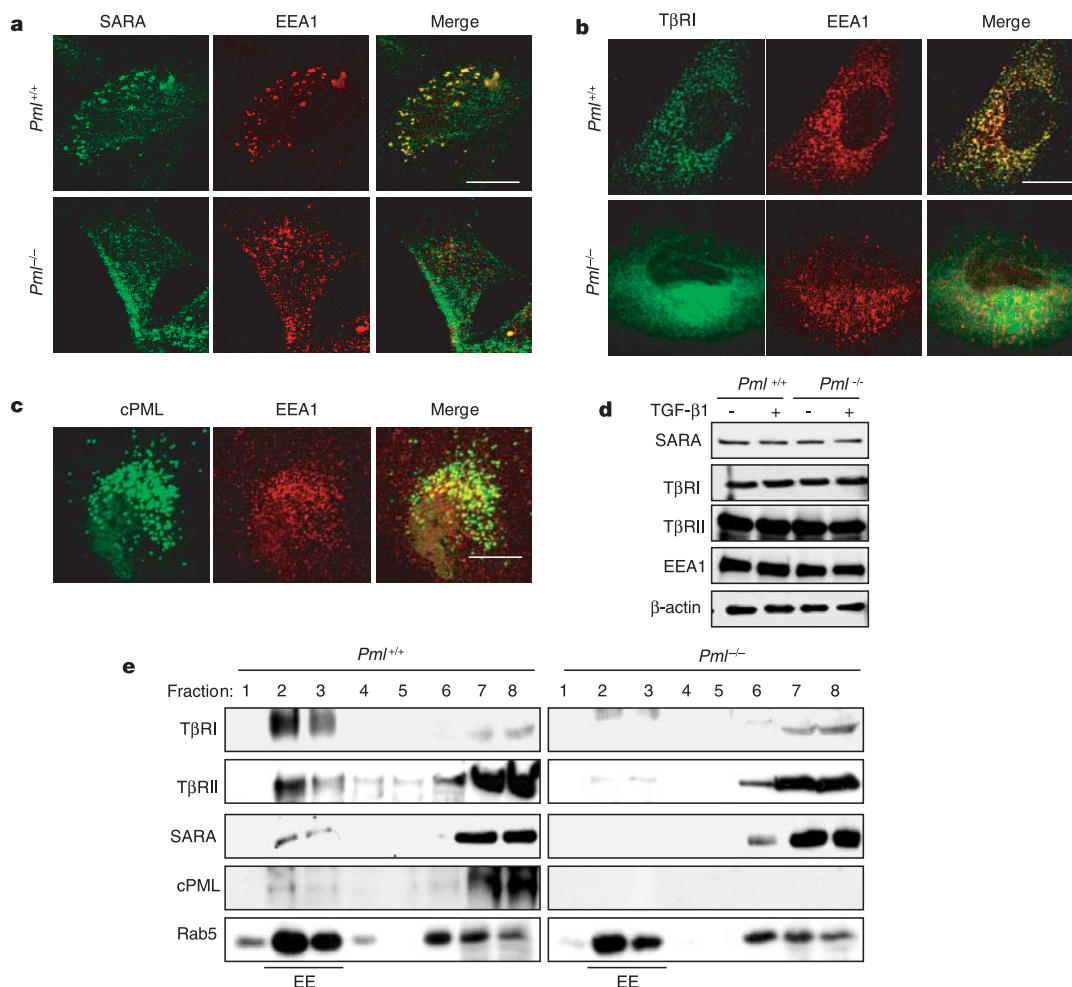
were transfected with the indicated plasmids for 48 h and harvested for Co-IP analysis. **g**, COS-1 cells were transfected with GFP-Smad3, with or without SARA and Xpress-cPML for 48 h, treated with or without TGF- $\beta$ 1 for 1 h and fixed for immunofluorescence. Nuclei were stained with DAPI. Scale bar, 10  $\mu$ m. **h**, *Pml*<sup>-/-</sup> immortalized MEFs were transfected with the indicated plasmids for 48 h and harvested for Co-IP analysis. **i**, HepG2 cells were co-transfected with Smad3 and the 3TP-luc reporter in the presence or absence of SARA and cPML for 3 h, treated with or without TGF- $\beta$ 1 for 30 h, and harvested for luciferase assay. Results are mean values  $\pm$  s.d. from a representative experiment performed in triplicate.

(amino acids 228–475), which no longer interacts with Smad3 (Fig. 1j, k; see also Supplementary Fig. 3a, b), rescued these phenotypes. These results therefore suggest that cPML may have an essential role in TGF- $\beta$  signalling.

As PML is readily induced after a number of cellular stresses, including oncogenic transformation<sup>10</sup>, and exposure to cytokines (for example, IFN; ref. 8), we examined whether TGF- $\beta$ 1 modulates cPML expression. Indeed, cPml expression was detected in wild-type MEFs, as well as in other cell lines, but not in *Pml*<sup>-/-</sup> MEFs (see Supplementary Fig. 1a, b). Notably, TGF- $\beta$ 1 increased expression of both cPml mRNA and protein (see Supplementary Fig. 1a, upper panel; and Supplementary Fig. 1c). Similarly, cPML was upregulated in human HepG2 cells after TGF- $\beta$ 1 treatment (see Supplementary Fig. 1a, lower panel). We next analysed the sub-cellular localization of cPML isoforms by confocal immunofluorescence. Using an anti-Pml antibody that recognizes the amino-terminal portion of Pml (shared by all Pml isoforms), the expected distinctive Pml-NB staining was observed, as well as a punctate cytoplasmic signal, which was specific as it was not detected in *Pml*<sup>-/-</sup> MEFs (see Supplementary Fig. 1d). Ectopic expression of cPML in *Pml*<sup>-/-</sup> MEFs produced a similar cytosolic punctate distribution, and cPML has a similar relative molecular mass (*M<sub>r</sub>*) to the mouse cPml isoform in western blot analysis (see

Supplementary Fig. 1d, f). Furthermore, western blot analysis confirmed that endogenous cPml, which, like human cPML, has an *M<sub>r</sub>* of around 75,000, was only detected in the cytoplasmic fraction, and not in the nuclear fraction or in *Pml*<sup>-/-</sup> MEFs (see Fig. 4d; and see also Supplementary Fig. 1e). Thus, cPML expression is induced by TGF- $\beta$ 1 and localizes in discrete cytosolic regions.

As *Pml* inactivation impairs TGF- $\beta$ 1-induced gene expression and the tumour-suppressive response, we sought to determine which step in the transduction of the TGF- $\beta$ 1 signal is affected as a consequence of *Pml* loss. At first, Smad2 and Smad3 phosphorylation were examined in wild-type and *Pml*<sup>-/-</sup> MEFs. TGF- $\beta$ 1-induced Smad2 and Smad3 phosphorylation was markedly impaired in *Pml*<sup>-/-</sup> MEFs (see Supplementary Fig. 2a–d). Intriguingly, overexpression of cPML enhanced TGF- $\beta$ 1-induced Smad2 and Smad3 phosphorylation in HepG2 cells (see Supplementary Fig. 2e), suggesting that cPML levels can have a critical function in the modulation of TGF- $\beta$ 1-induced Smad phosphorylation. Furthermore, TGF- $\beta$ 1 triggered Smad3 nuclear translocation in wild-type, but not in *Pml*<sup>-/-</sup> MEFs (see Supplementary Fig. 2f, g). Ectopic expression of cPML in *Pml*<sup>-/-</sup> MEFs fully restored TGF- $\beta$ 1-induced Smad3 nuclear transport (see Supplementary Fig. 2h). These observations suggest that cPml alone is sufficient to rescue the TGF- $\beta$  response after *Pml* loss.



**Figure 3** PML is required for TβRI/II and SARA localization in the early endosome. **a**, Wild-type or *Pml*<sup>-/-</sup> MEFs were harvested and fixed for immunofluorescence microscopy. **b**, MEFs were transfected with a constitutively active haemagglutinin-tagged TβRI (HA-TβRI; T204D) for 30 h. Cells were then kept at 4 °C for 15 min, incubated for 30 min at 37 °C and fixed for immunofluorescence. **c**, Wild-type MEFs were transfected with cPML plasmid for 48 h and fixed for immunofluorescence microscopy. Scale bar,

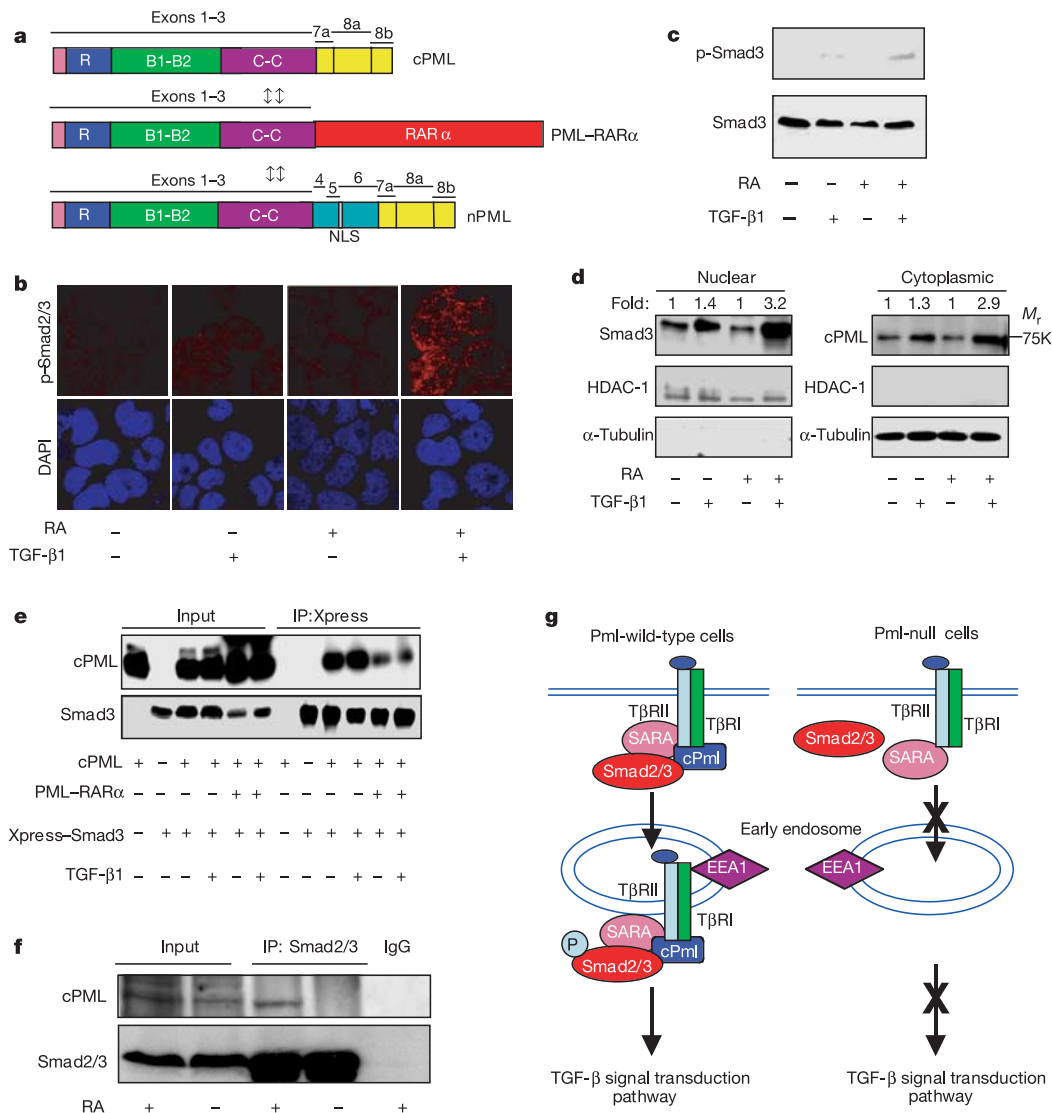
20 μm. **d**, MEFs were either left untreated or were treated with TGF- $\beta$  for 1 h and cells harvested for western blot analysis. **e**, MEFs were treated with TGF- $\beta$  at 4 °C for 15 min. Cells were then incubated for 30 min at 37 °C and harvested for sucrose gradient floatation, followed by western blot analysis (see Methods section). EE indicates early endosome/Rab5-enriched fractions.

We next investigated the mechanisms by which cPML modulates TGF- $\beta$ -signalling. Co-immunoprecipitation assays demonstrated that exogenous cPML interacted with exogenous Smad2 and Smad3 irrespective of TGF- $\beta$ 1 treatment (Fig. 2a, b). Smad3 interacted with the two B-box domains of cPML (see Supplementary Fig. 3a, b), whereas cPML interacted with the Smad3 MH1 domain (see Supplementary Fig. 3c, d). Endogenous cPml also associated effectively with endogenous Smad2/3, but this association was lost after TGF- $\beta$ 1 treatment (Fig. 2c).

As SARA interacts with Smad2/3, we examined whether cPML also binds to SARA. Exogenous SARA was detected in the cPML immunocomplex both in the presence or absence of TGF- $\beta$ 1 (Fig. 2d). Similar to cPML and Smad2/3 association, interaction of endogenous SARA and cPML was readily observed at the steady state, but lost in the presence of TGF- $\beta$ 1 (Fig. 2e). Furthermore,

both SARA and cPML were detected in the Smad2/3 immunocomplex, suggesting that Smad2/3 can associate with both cPML and SARA (Fig. 2c). Next, we determined whether cPML would co-localize with SARA and/or Smad3. cPML, SARA and Smad3 also readily colocalized in cytosolic regions (Fig. 2g; see also Supplementary Fig. 4).

We next set out to analyse the function of cPML in the formation of the SARA/Smad2/3 complex. Notably, SARA and Smad2/3 no longer interacted in the absence of Pml (Fig. 2c, right panel), although the two proteins were expressed at comparable levels in wild-type and *Pml*<sup>-/-</sup> MEFs (Fig. 2c, e). Furthermore, SARA and Smad3 failed to colocalize in *Pml*<sup>-/-</sup> MEFs, whereas they colocalized in wild-type MEFs (see Supplementary Fig. 5). Add-back of cPML in *Pml*<sup>-/-</sup> MEFs fully restored the SARA and Smad2/3 interaction (Fig. 2h). To understand how cPML could mediate



**Figure 4** Aberrant TGF- $\beta$ 1 signalling in APL. **a**, Schematic representation of PML-RAR $\alpha$ , cPML and a nuclear PML (nPML) isoform and their interaction interfaces (arrows). The fusion protein physically interacts with cPML and nPML through the RBCC domain. R, RING finger domain; B1-B2, B box domains; C-C, coiled coil domain; NLS, nuclear localization signal. NB4 cells were pretreated with vehicle or 1  $\mu$ M RA for 24 h, followed by 1 h treatment with TGF- $\beta$ 1. Cells were then fixed for immunofluorescence microscopy with anti-phospho Smad2/3 antibody (**b**), or western blot analysis with anti-phospho-Smad3 antibody (**c**). Cells were also harvested for preparation of nuclear and cytoplasmic fractions, followed by western blot analysis (**d**). **e**, 293T cells were transfected with the

indicated plasmids for 48 h, treated with or without TGF- $\beta$ 1 for 1 h, and harvested for Co-IP analysis. **f**, NB4 cells were treated with vehicle or RA for 48 h and cells were harvested for Co-IP analysis. **g**, Model for the role of cPML in TGF- $\beta$ 1 signalling. In wild-type cells, cPML complexes with SARA and Smad2/3. This is essential for SARA and T $\beta$ RI/II localization in the early endosome and in turn allows Smad2/3 phosphorylation and nuclear translocation, after TGF- $\beta$ 1 induction, to occur. In contrast, in MEFs devoid of PML, the interaction of SARA with Smad2/3 as well as the localization of SARA and T $\beta$ RI/II in the early endosome are markedly impaired, resulting in defective signal transduction.



SARA and Smad2/3 interaction, we used a SARA mutant lacking the Smad binding domain (SARA  $\Delta$ SBD)<sup>13</sup>. As expected, SARA  $\Delta$ SBD did not interact with Smad3. Notably, in the presence of ectopically expressed cPML, Smad3 associated with SARA  $\Delta$ SBD, and both cPML and SARA  $\Delta$ SBD were detected in the Smad3 immunocomplexes (Fig. 2f), suggesting that cPML may function as a bridging factor between SARA and Smad3. Therefore, although SARA and Smad2/3 can interact even in the absence of cPML *in vitro*<sup>13</sup>, cPML is required for this interaction *in vivo*. As SARA regulates TGF- $\beta$ 1-mediated gene expression<sup>13</sup>, we assessed whether cPML would affect the ability of SARA to regulate TGF- $\beta$ 1-dependent transcription. Indeed, when coexpressed, cPML and SARA markedly potentiated TGF- $\beta$ 1-induced gene expression (Fig. 2i). Thus, cPML is essential for the formation of a stable and functional SARA/Smad2/3 complex.

Several reports have demonstrated that TGF- $\beta$  type I/II receptors (T $\beta$ RI/II) are internalized into the early endosome through the clathrin-dependent pathway<sup>14–16</sup>. However, instead of targeting TGF- $\beta$  receptor for degradation, the clathrin/early endosome pathway is critical for TGF- $\beta$  signal transduction. SARA, on the other hand, is located in the early endosome and facilitates T $\beta$ RI/II transport into this subcellular compartment<sup>14,16</sup>. Immunofluorescence microscopy revealed the colocalization of SARA, T $\beta$ RI and cPML with early endosome antigen-1 (EEA1), an early endosome marker, in wild-type MEFs (Fig. 3a–c; see also Supplementary Fig. 6, 7). By contrast, in *Pml*<sup>−/−</sup> MEFs, SARA localization was aberrant and colocalization of SARA and EEA1 was markedly reduced (Fig. 3a). T $\beta$ RI localization and colocalization with EEA1 was similarly impaired in *Pml*<sup>−/−</sup> MEFs (Fig. 3b; see also Supplementary Fig. 6). This aberrant SARA and T $\beta$ RI localization pattern was not due to changes in protein amounts in *Pml*<sup>−/−</sup> MEFs (Fig. 3d).

To further corroborate our observations, we performed sucrose flotation gradients to isolate the endosome compartment enriched for Rab5 (an early endosome marker). In wild-type MEFs, SARA, T $\beta$ RI/II and cPML accumulated in the Rab5–early-endosome-enriched fractions (Fig. 3e; left panel, fractions 2, 3). By contrast, the amount of SARA and T $\beta$ RI/II in these fractions was markedly reduced in *Pml*<sup>−/−</sup> MEFs (Fig. 3e, right panel, fractions 2, 3). As T $\beta$ RI/II localization to the early endosome was impaired in *Pml*<sup>−/−</sup> MEFs, we determined whether cPML interacts with T $\beta$ RI and/or T $\beta$ RII. Co-immunoprecipitation experiments demonstrated that cPML interacts with both T $\beta$ RI and T $\beta$ RII (see Supplementary Fig. 8), and TGF- $\beta$  treatment resulted in the dissociation of cPML and the T $\beta$ RI/II complex. Together, these results suggest that cPML is present in the T $\beta$ RI/T $\beta$ RII/SARA/Smad complex and that it facilitates the localization of this complex into the early endosome. By contrast, epidermal growth factor receptor transport through the early endosome was not impaired in the absence of *Pml* (data not shown).

APL blasts are poorly responsive to TGF- $\beta$ <sup>17–19</sup>. However, the molecular mechanisms underlying this phenomenon are unknown. In APL leukaemic cells, PML–RAR $\alpha$  is found both in the nuclear and cytosolic fractions<sup>9</sup> where it can heterodimerize with both cPML and nuclear PML isoforms (nPML) through the RBCC PML domain, fused N-terminally to the RAR $\alpha$  moiety (ref. 8; Fig. 4a; and data not shown). Because of the essential role of Pml in TGF- $\beta$  signalling, we hypothesized that the presence of PML–RAR $\alpha$  in the APL blasts could functionally phenocopy Pml inactivation. To test this, we used a paradigmatic APL cell line, NB4, which expresses the PML–RAR $\alpha$  fusion protein and is known to respond to retinoic acid (RA) and As<sub>2</sub>O<sub>3</sub> (ref. 20). Similarly to the observations in *Pml*<sup>−/−</sup> MEFs, TGF- $\beta$ 1-induced Smad2/3 phosphorylation and Smad3 nuclear translocation was defective in NB4 cells (Fig. 4b–d). TGF- $\beta$ 1-dependent induction of cPML was also markedly impaired (Fig. 4d). RA treatment induced degradation of PML–RAR $\alpha$  and restored TGF- $\beta$ 1-dependent Smad2/3 phosphorylation and Smad3 nuclear translocation, as well as cPML

expression (Fig. 4b–d and data not shown). As<sub>2</sub>O<sub>3</sub> similarly restored TGF- $\beta$ 1-dependent Smad2/3 phosphorylation and nuclear translocation (data not shown). Importantly, PML–RAR $\alpha$  could function as a dominant-negative cPML mutant to interrupt the association of cPML with Smad3 either in the presence or absence of TGF- $\beta$ 1 (Fig. 4e). RA treatment, which triggers PML–RAR $\alpha$  degradation<sup>20</sup>, rescued the interaction between cPML and Smad2/3 (Fig. 4f). These data therefore demonstrate that TGF- $\beta$ 1 signalling is impaired in APL blasts, similar to *Pml*<sup>−/−</sup> cells, and support the notion that the PML–RAR $\alpha$  oncoprotein antagonizes cPML function, in turn resulting in TGF- $\beta$  unresponsiveness.

We therefore attribute a key biological function to cytosolic PML isoforms. It is important to note, however, that as nPML shuttles between the nucleus and the cytoplasm after viral infection<sup>21</sup>, it is possible that nPML isoforms could function in the cytosol under physiological conditions. In agreement with this notion, nPML isoforms, when artificially targeted to the cytoplasm, are also capable of modulating TGF- $\beta$  signalling. From an evolutionary stand point, it will be of interest to assess whether cPML came before nPML, and if the PML–NB has only more recently evolved to provide a more specialized function. Furthermore, we have identified PML as a key regulator of TGF- $\beta$  signalling and function. We propose a model to account for the role of cPML in this process (Fig. 4g). It has been recently reported that p53, which, like PML is dispensable for embryogenesis, is also required for TGF- $\beta$ /Smad-dependent transcriptional activity and biological functions<sup>22</sup>. This suggests that PML and p53 may participate in a more restricted regulatory network for the modulation of TGF- $\beta$ /Smad-dependent functions in response to oncogenic and physiological stress. In this respect, preliminary data indicate that *Pml*-null mice display a defect in the TGF- $\beta$ -mediated wound healing process (H.-K.L., S.B. and P.P.P., unpublished observations). Finally, we attribute a discrete oncogenic function to cytosolic PML–RAR $\alpha$  in APL pathogenesis through its ability to prevent cPML from participating in the transduction of the TGF- $\beta$ 1 pathway. The profound impairment in TGF- $\beta$ -dependent tumour-suppressive activities observed in *Pml*-null cells is in agreement with the frequent loss of PML in tumour types, such as colon cancer<sup>23</sup>, where an impairment in the TGF- $\beta$ 1 pathway (for example, through Smad mutations<sup>1,24,25</sup>) was previously documented. □

## Methods

### Mice, cell culture and reagents

MEFs from wild-type and *Pml*<sup>−/−</sup> mice were prepared as previously described<sup>26</sup>. 293T cells, COS-1 cells (from ATCC) and HepG2 cells (a gift from C. Chang) were cultured in DMEM containing 10% fetal bovine serum (FBS). Xpress–Smad3, Xpress–Smad3 (amino acids 1–270), Xpress–Smad3 (amino acids 170–425) and Xpress–Smad3 (amino acids 270–425), were all cloned after polymerase chain reaction (PCR) amplification of Smad3, Smad3 (amino acids 1–270), Smad3 (amino acids 170–425) and Smad3 (amino acids 270–425) using Flag–Smad3<sup>27</sup> as a template and insertion into pCDNA4/HisMax–TOPO vector (Invitrogen). To clone GFP–Smad3, Smad3 was amplified by PCR using Flag–Smad3 as a template and inserted into pCDNA3.1/NT–GFP–TOPO vector (Invitrogen). Flag–Smad2 and Flag–SARA were gifts from J. Massagué. pCDNA3–cPML (PML3 3–7; ref. 6) was from P.G. Pelicci. Xpress–cPML, Xpress–cPML (amino acids 98–475) and Xpress–cPML (amino acids 228–475), were cloned after PCR amplification of cPML, cPML 98–475, and cPML 228–475 and pCDNA3–cPML used as a template before insertion of products into pCDNA4/HisMax–TOPO vector (Invitrogen). Flag–SARA  $\Delta$ SBD and PML4  $\Delta$ NLS were gifts from J.L. Wrana and C.S. Chang, respectively. TGF- $\beta$ 1 was purchased from Calbiochem.

### Immunoprecipitation, immunoblotting and immunofluorescence

Immunoprecipitation (IP), immunoblotting (IB) and immunofluorescence (IF) were performed as described with some modifications<sup>28</sup>. For endogenous protein–protein interaction, cells were lysed with RIPA buffer (1 × PBS, 0.1% SDS, 0.5% sodium deoxycholate, 0.5% NP-40, 5% glycerol and protease inhibitor cocktail, Roche). The following antibodies were used for IP, IB or IF: mouse anti-mouse Pml (IP: 1:250; IB/IF: 1:500, S.W. Lowe), mouse anti-human PML (IP: 1:100, Santa Cruz), rabbit anti-human PML (IB: 1:5,000, IF: 1:1,000, C.S. Chang), anti-phospho-Smad2 (IB: 1:1,000, IF: 1:250, Cell Signaling), anti-Smad2/3 (IP: 1:100; IB: 1:500, BD Transduction Lab.), anti-phospho-Smad3 (IB: 1:500, M. Reiss), anti-Smad3 (IB: 1:500, IF: 1:250, Zymed), anti-phospho-Smad23 (IF: 1:50, Santa Cruz), anti-SARA (IP: 1:100; IB/IF: 1:500, Santa Cruz), anti-T $\beta$ RI (IB: 1:500, IF: 1:250, Santa Cruz), anti-T $\beta$ RII (IB: 1:1,000, Santa Cruz), anti-rab5

(IB: 1:1,000, BD Transduction Lab.), anti-Xpress (IP/IF: 1:500; IB: 1:5,000, Invitrogen), anti-Flag (IP: 1:500; IB: 1:2,000, Sigma), anti-HDAC-1 (IB: 1:1,000), anti-E2F1 (IB: 1:500, Santa Cruz), anti- $\alpha$ -tubulin (IB: 1:1,000), anti- $\beta$ -actin (IB: 1:1,000, Sigma), anti-p21 (IB: 1:500, Pharmingen), anti-p15 (IB: 1:500, Santa Cruz) and anti-EEA1 (IF: 1:250, BD Transduction Lab.).

#### Apoptosis, cell growth and senescence assay

B cells isolated from 8–12-week-old wild-type and *Pml*<sup>-/-</sup> mice were treated with vehicle or TGF- $\beta$ 1 for 30 h. Cells were fixed, and the TUNEL assay was performed using the *in situ* cell death detection kit (Roche) according to the manufacturer's instructions. For growth assays, after TGF- $\beta$ 1 treatment, cells were harvested and stained by trypan blue at different days of TGF- $\beta$ 1 treatment, and numbers of viable cells were directly counted by light microscopy. A thymidine incorporation assay was performed as previously described<sup>29</sup>. For senescence assays, wild-type or *Pml*<sup>-/-</sup> MEFs were either left untreated or treated with TGF- $\beta$ 1 for 6 days, and cells were fixed and stained with  $\beta$ -galactosidase ( $\beta$ -gal) using the senescence detection kit (Oncogene).

#### Sucrose flotation gradient

Early-endosome-containing fractions were prepared as described<sup>30</sup>, with some modifications. In brief, MEFs (6–8 near-confluent 10-mm dishes) cultures were washed twice with ice-cold PBS, scraped in PBS and centrifuged at 1,200 g. After addition of 5 ml homogenization buffer (250 mM sucrose, 3 mM imidazole at pH 7.4, 0.5 mM EDTA), cell pellets were put on ice for 5 min, centrifuged and resuspended in 0.5 ml homogenization buffer containing the protease inhibitor cocktail (Roche). Cells were then homogenized at 4 °C by seven passages through a 25G 5/8 needle fitted onto a 1 ml plastic syringe. Homogenates were centrifuged for 10 min at 2,000 g at 4 °C and the post-nuclear supernatants (PNSs) were collected. The PNS fraction was then brought to 40.6% sucrose using a stock solution (62% sucrose, 3 mM imidazole at pH 7.4) (final volume 1.1 ml) and loaded at the bottom of an SW 60 centrifugation tube. A gradient consisting of three steps was then poured (1.5 ml of 35% sucrose, 3 mM imidazole at pH 7.4; 1 ml 30% sucrose, 3 mM imidazole at pH 7.4; 0.5 ml homogenization buffer). The gradient was centrifuged at 485,000 g for 90 min using an SW60 rotor. Eight 0.5 ml fractions were collected from the top of the tube, and a portion (40  $\mu$ l) of each was subjected to SDS-PAGE analysis, followed by western blot analysis.

#### Transcriptional assays

MEFs and HepG2 cells were transfected by using superfect (Qiagen) for 3 h, and cells were treated with vehicle or 100 pM TGF- $\beta$ 1 for 30 h. The luciferase activity was measured by using the dual luciferase system (Promega). The results were normalized over renilla activity.

Received 27 March; accepted 25 June 2004; doi:10.1038/nature02783.

- Derynck, R., Akhurst, R. J. & Balmann, A. TGF- $\beta$  signaling in tumor suppression and cancer progression. *Nature Genet.* **29**, 117–129 (2001).
- Siegel, P. M. & Massague, J. Cytostatic and apoptotic actions of TGF- $\beta$  in homeostasis and cancer. *Nature Rev. Cancer* **3**, 807–820 (2003).
- Attisano, L. & Wrana, J. L. Signal transduction by the TGF- $\beta$  superfamily. *Science* **296**, 1646–1647 (2002).
- Kim, S. J. & Letterio, J. Transforming growth factor- $\beta$  signaling in normal and malignant hematopoiesis. *Leukemia* **17**, 1731–1737 (2003).
- Jensen, K., Shiels, C. & Freemont, P. S. PML protein isoforms and the RBCC/TRIM motif. *Oncogene* **20**, 7223–7233 (2001).
- Fagioli, M. *et al.* Alternative splicing of PML transcripts predicts coexpression of several carboxy-terminally different protein isoforms. *Oncogene* **7**, 1083–1091 (1992).
- Melnick, A. & Licht, J. D. Deconstructing a disease: RAR $\alpha$ , its fusion partners, and their roles in the pathogenesis of acute promyelocytic leukemia. *Blood* **93**, 3167–3215 (1999).
- Salomoni, P. & Pandolfi, P. P. The role of PML in tumor suppression. *Cell* **108**, 165–170 (2002).
- Kastner, P. *et al.* Structure, localization and transcriptional properties of two classes of retinoic acid receptor  $\alpha$  fusion proteins in acute promyelocytic leukemia (APL): structural similarities with a new family of oncoproteins. *EMBO J.* **11**, 629–642 (1992).
- Ferbeyre, G. *et al.* PML is induced by oncogenic ras and promotes premature senescence. *Genes Dev.* **14**, 2015–2027 (2000).
- Pearson, M. *et al.* PML regulates p53 acetylation and premature senescence induced by oncogenic Ras. *Nature* **406**, 207–210 (2000).
- Katakura, Y., Nakata, E., Miura, T. & Shirahata, S. Transforming growth factor  $\beta$  triggers two independent-senescence programs in cancer cells. *Biochem. Biophys. Res. Commun.* **255**, 110–115 (1999).
- Tsukazaki, T., Chiang, T. A., Davison, A. F., Attisano, L. & Wrana, J. L. SARA, a FYVE domain protein that recruits Smad2 to the TGF $\beta$  receptor. *Cell* **95**, 779–791 (1998).
- Di Guglielmo, G. M., Le Roy, C., Goodfellow, A. F. & Wrana, J. L. Distinct endocytic pathways regulate TGF- $\beta$  receptor signalling and turnover. *Nature Cell Biol.* **5**, 410–421 (2003).
- Panopoulou, E. *et al.* Early endosomal regulation of Smad-dependent signaling in endothelial cells. *J. Biol. Chem.* **277**, 18046–18052 (2002).
- Hayes, S., Chawla, A. & Corvera, S. TGF  $\beta$  receptor internalization into EEA1-enriched early endosomes: role in signaling to Smad2. *J. Cell. Biol.* **158**, 1239–1249 (2002).
- Testa, U. *et al.* PML/RAR  $\alpha$ +U937 mutant and NB4 cell lines: retinoic acid restores the monocytic differentiation response to vitamin D3. *Cancer Res.* **54**, 4508–4515 (1994).
- Grignani, F. *et al.* The acute promyelocytic leukemia-specific PML-RAR  $\alpha$  fusion protein inhibits differentiation and promotes survival of myeloid precursor cells. *Cell* **74**, 423–431 (1993).
- Testa, U. *et al.* Transforming growth factor- $\beta$  potentiates vitamin D3-induced terminal monocytic differentiation of human leukemic cell lines. *J. Immunol.* **150**, 2418–2430 (1993).
- Zhu, J., Chen, Z., Lallemand-Breitenbach, V. & de Thé, H. How acute promyelocytic leukaemia revived arsenic. *Nature Rev. Cancer* **2**, 705–713 (2002).

- Turelli, P. *et al.* Cytoplasmic recruitment of INI1 and PML on incoming HIV preintegration complexes: interference with early steps of viral replication. *Mol. Cell* **7**, 1245–1254 (2001).
- Cordenonsi, M. *et al.* Links between tumor suppressors: p53 is required for TGF- $\beta$  gene responses by cooperating with Smads. *Cell* **113**, 301–314 (2003).
- Gurrieri, C. *et al.* Loss of the tumor suppressor PML in human cancers of multiple histologic origins. *J. Natl Cancer Inst.* **96**, 269–279 (2004).
- Howe, J. R. *et al.* Mutations in the SMAD4/DPC4 gene in juvenile polyposis. *Science* **280**, 1086–1088 (1998).
- Miyaki, M. *et al.* Higher frequency of Smad4 gene mutation in human colorectal cancer with distant metastasis. *Oncogene* **18**, 3098–3103 (1999).
- Wang, Z. G. *et al.* Role of PML in cell growth and the retinoic acid pathway. *Science* **279**, 1547–1551 (1998).
- Kang, H. Y. *et al.* From transforming growth factor- $\beta$  signaling to androgen action: identification of Smad3 as an androgen receptor coregulator in prostate cancer cells. *Proc. Natl Acad. Sci. USA* **98**, 3018–3023 (2001).
- Lin, H. K., Wang, L., Hu, Y. C., Altuwajri, S. & Chang, C. Phosphorylation-dependent ubiquitylation and degradation of androgen receptor by Akt require Mdm2 E3 ligase. *EMBO J.* **21**, 4037–4048 (2002).
- Kurokawa, M. *et al.* The oncoprotein Evi-1 represses TGF- $\beta$  signalling by inhibiting Smad3. *Nature* **394**, 92–96 (1998).
- Escola, J. M., Grivel, J. C., Chavrier, P. & Gorvel, J. P. Different endocytic compartments are involved in the tight association of class II molecules with processed hen egg lysozyme and ribonuclease A in B cells. *J. Cell Sci.* **108**, 2337–2345 (1995).

**Supplementary Information** accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We are grateful to C. Chang, K. S. Chang, H. Y. Kang, Y. B. Kang, S. W. Lowe, P. G. Pelicci, M. Reiss, T. Sollner and J. L. Wrana for reagents and advice. Special thanks to all the members of the Pandolfi laboratory for comments and discussion. We also thank T. Merghoub and I. Guernah for technical help, R. Hobbs and L. DiSantis for critical reading and preparation of the manuscript. This work was supported by National Institutes of Health grants to P.P.P.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to P.P.P. (p-pandolfi@ski.mskcc.org).

## Induction of DNA methylation and gene silencing by short interfering RNAs in human cells

Hiroaki Kawasaki & Kazunari Taira

Department of Chemistry and Biotechnology, School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan, and Gene Function Research Center, National Institute of Advanced Industrial Science and Technology (AIST), Central 4, 1-1-1 Higashi, Tsukuba Science City 305-8562, Japan

Double-stranded RNAs (dsRNAs) induce post-transcriptional gene silencing in several species of animal and plant<sup>1–2</sup>. In plants, dsRNAs targeted to CpG islands within a promoter can also induce RNA-directed DNA methylation<sup>3–8</sup>; however, it remains unclear whether gene silencing mediated by DNA methylation can be induced by dsRNAs in mammalian cells. Here, we demonstrate that short interfering RNAs (siRNAs; 21–25-nucleotide RNA molecules) induce DNA methylation and histone H3 methylation in human cells. Synthetic siRNAs targeted to CpG islands of an E-cadherin promoter induced significant DNA methylation and histone H3 lysine 9 methylation in both MCF-7 and normal mammary epithelial cells. As a result, these siRNAs repressed expression of the E-cadherin gene at the transcriptional level. In addition, disrupting the expression of either one of two DNA methyltransferases (*DNMT1* or *DNMT3B*) by specific siRNAs abolished the siRNA-mediated methylation of DNA. Moreover, vector-based siRNAs targeted to the *erbB2* (also known as *HER2*) promoter also induced DNA methylation in MCF-7 cells. Thus, siRNAs targeted to CpG

islands within the promoter of a specific gene can induce transcriptional gene silencing by means of DNA-methyltransferase-dependent methylation of DNA in human cells, and might have potential as a new type of gene therapeutic agent.

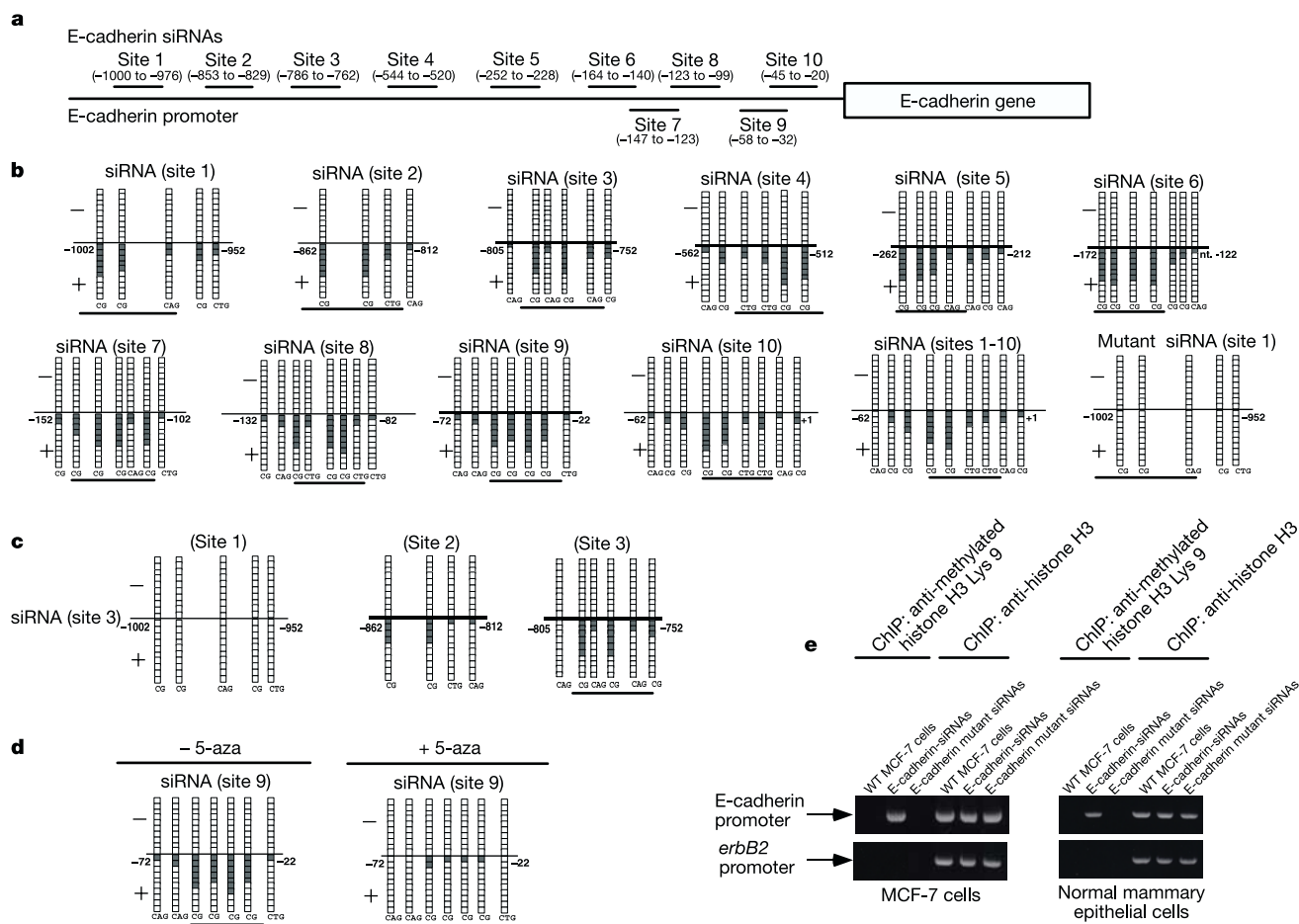
Double-stranded RNAs induce RNA interference (RNAi)-mediated post-transcriptional gene silencing in animals and plants<sup>1–2</sup>. In this system, siRNAs are generated by the RNase III Dicer enzyme and are incorporated into the RNAi-induced silencing complex (RISC)<sup>9,10</sup>. The siRNA–RISC complex then promotes degradation of cytoplasmic messenger RNAs<sup>11–16</sup>.

To determine whether synthetic siRNAs might induce DNA methylation in mammalian cells, we synthesized siRNAs targeted to CpG islands of the E-cadherin promoter (E-cadherin-siRNAs), as E-cadherin can be silenced by aberrant methylation of the promoter in several lines of tumour cells<sup>17–19</sup>. Moreover, as CpG sites in the E-cadherin promoter are unmethylated in MCF-7 cells<sup>17,18</sup>, we chose MCF-7 cells for this study. We selected ten CpG sites (sites 1–10) in the E-cadherin promoter as targets of siRNAs (Fig. 1a). Each E-cadherin-siRNA (100 nM) was introduced into MCF-7 cells and, 96 h later, total DNA was collected and genomic DNA was isolated. We then performed bisulphite sequencing to determine the methylation status of the E-cadherin promoter. We treated genomic DNA with bisulphite using a CpGenome DNA-modification kit

(InterGen). We amplified the bisulphite-modified E-cadherin promoter by polymerase chain reaction (PCR) and cloned the product of PCR using a TA cloning kit (Clontech). We picked ten independent colonies for each target site and analysed the cloned sequences by direct sequencing.

As shown in Fig. 1b, we found methylated DNA in MCF-7 cells that harboured each respective E-cadherin-siRNA (sites 1–10) and cells harbouring all E-cadherin-siRNAs (sites 1–10; the number of grey squares indicates the number of methylated cytosines). By contrast, a mutant form of the E-cadherin-siRNA directed against site 1 that had eight point mutations in both its sense and antisense strand failed to induce methylation. Moreover, a combination of all E-cadherin-siRNAs together failed to induce DNA methylation within a non-targeted *erbB2* promoter (Supplementary Fig. S1), highlighting the specificity of the E-cadherin-siRNAs. We next examined the possibility of methylation beyond the target site in the absence of RNA-directed RNA polymerase (RdRP) in mammalian cells. Notably, siRNAs targeted to site 3 of the E-cadherin promoter induced methylation not only at that site but also at the adjacent site (site 2), at least to some extent (Fig. 1c). However, site 1, located ~200 nucleotides upstream of site 3, was unmethylated.

Aberrant methylation is often detected in cancer cells. Therefore, we next examined the induction of DNA methylation by siRNAs in



**Figure 1** DNA methylation of the E-cadherin promoter by siRNAs in MCF-7 and normal breast epithelial cells. **a**, The ten target sites for siRNAs in the E-cadherin promoter. Nucleotides are indicated in parentheses. **b**, Induction of DNA methylation within the E-cadherin promoter in MCF-7 cells by siRNAs, as determined by bisulphite sequencing (see text for details). The number (out of a possible ten) and position of methylated cytosines are indicated by grey squares; underlining indicates the target sequences of siRNAs. **c**, Details of extended methylated regions within the E-cadherin promoter when siRNA was targeted to site 3 within the E-cadherin promoter in MCF-7 cells, as determined

by bisulphite sequencing. **d**, Relative levels of DNA methylation of the E-cadherin promoter at site 9 in MCF7 cells in the presence or absence of 5-aza, as determined by bisulphite sequencing. **e**, Detection of methylated histone associated with the E-cadherin promoter in cells that were treated with a combination of E-cadherin siRNAs (site 1–10). Methylated histone H3 at lysine 9 was detected using a chromatin immunoprecipitation (ChIP) assay with specific antibodies for methylated histone H3 at lysine 9. Histone H3 antibody is a positive control. WT, wild type.



normal breast epithelial cells. We detected methylation of CpG sites in normal breast epithelial cells that had been treated with E-cadherin-siRNAs (testing methylation at sites 6 and 10; Supplementary Fig. S2). Furthermore, the extent of methylation in MCF-7 cells that had been treated with a combination of all ten E-cadherin-siRNAs was reduced in the presence of 1  $\mu$ M 5-aza-2'-deoxycytidine (5-aza), an inhibitor of DNA methylation (Fig. 1d).

As methylation of histone H3 at lysine 9 is induced by RNAi in plants, fission yeast and *Drosophila*<sup>8,20,21</sup>, we examined whether our set of siRNAs could induce its methylation in mammalian cells using a chromatin immunoprecipitation assay with specific antibodies for methylated histone H3 at lysine 9. Each cell lysate from MCF-7 cells and normal mammary epithelial cells in the presence or absence of the E-cadherin-siRNAs (sites 1–10) was incubated with specific antibodies for methylated histone H3 at lysine 9 or antibodies for histone H3, and the DNA interacting with the methylated histone H3 was immunoprecipitated. The precipitated DNA was then amplified using specific primers for the E-cadherin promoter. As shown in Fig. 1e, the precipitated DNA from the cell lysate of E-cadherin-siRNA-treated MCF-7 cells produced a band corresponding to the E-cadherin promoter region. By contrast, we did not detect the corresponding band when the same procedure was performed in the absence of E-cadherin-siRNAs or in the presence of mutant siRNAs. Similar results were obtained using cell lysate from normal mammary epithelial cells. In addition, it has been reported that an RNA component is involved in maintenance or stabilization of a higher-order structure at pericentric heterochromatin in mammalian cells<sup>22</sup>.

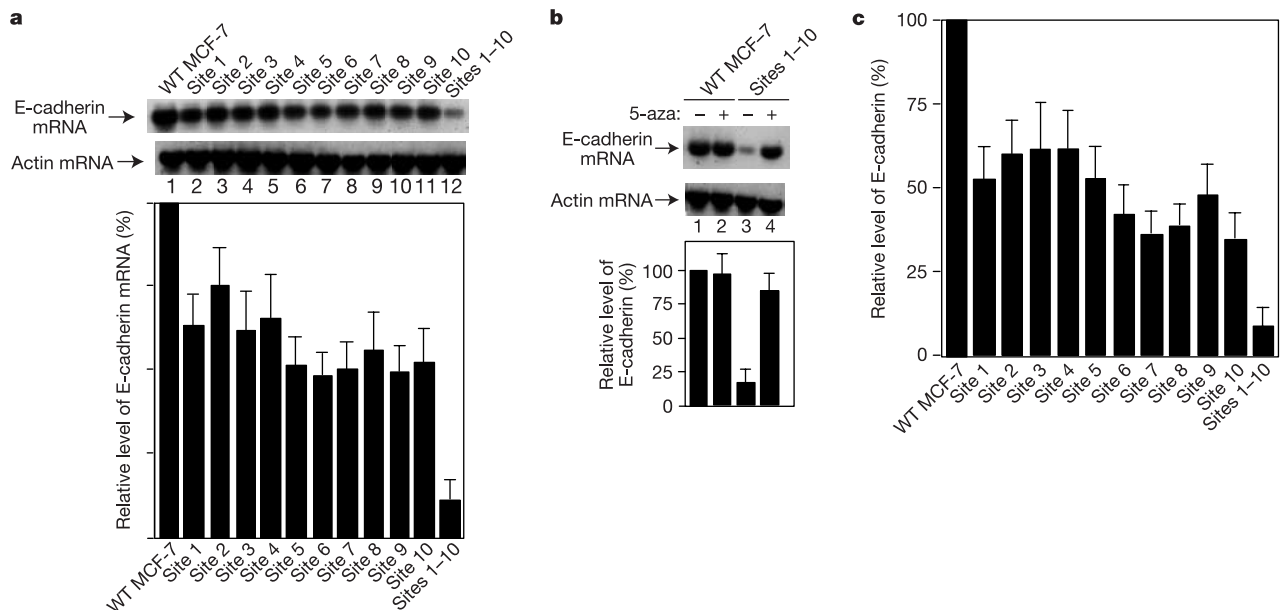
Taken together, these results suggest that siRNAs targeted to the E-cadherin promoter can induce not only DNA methylation but also histone H3 methylation at lysine 9 in human cancer cells and normal cells, and that the effects of siRNAs can spread to nearby regions even in the absence of RdRP.

To examine whether the induction of DNA methylation by E-cadherin-siRNAs was correlated with expression of the E-cadherin gene, we performed northern blotting analysis with a probe specific for E-cadherin mRNA. As shown in Fig. 2a, the level of E-cadherin

mRNA expression in each line of MCF-7 cells harbouring a specific E-cadherin-siRNA was lower than in wild-type MCF-7 cells (lanes 1–11). Notably, the level of E-cadherin mRNA expression in MCF-7 cells that had been treated with a combination of all E-cadherin-siRNAs (lane 12) was significantly lower than in MCF-7 cells transfected with a single siRNA, clearly demonstrating the additive effects of the siRNAs. Treatment with 5-aza almost completely abolished the effect of the E-cadherin-siRNAs (Fig. 2b, lane 4). Western blot analysis with an E-cadherin-specific antibody also revealed similar reductions in the amount of E-cadherin (Fig. 2c). Thus, siRNAs targeted to the E-cadherin promoter act as gene silencers at the transcriptional level, and induction by siRNAs of DNA methylation within the E-cadherin promoter is inversely correlated with the level of gene expression and is additive.

DNA methyltransferases (DNMTs) are responsible for all DNA methylation in the cell<sup>23–25</sup>. To determine whether DNMTs participate in siRNA-mediated DNA methylation in human cells, we tried to suppress the expression of genes for DNMTs using siRNAs targeted to the respective mRNAs. We synthesized three separate siRNAs that targeted to *DNMT1*, *DNMT2* and *DNMT3B* mRNAs, respectively. As controls, we used mutant *DNMT1*-, *DNMT2*- and *DNMT3B*-siRNAs with four point mutations in both the sense and antisense strand. We introduced *DNMT*-siRNA or mutant *DNMT1*-siRNA at 100 nM into MCF-7 cells using Oligofectamine, and examined the amount of DNMTs by western blotting. Amounts of *DNMT1*, *DNMT2* and *DNMT3B* in MCF-7 cells transfected with *DNMT1*-siRNA, *DNMT2*-siRNA or *DNMT3B*-siRNA, respectively, but not the corresponding mutant *DNMT*-siRNAs, were significantly lower than those in wild-type MCF-7 cells (Fig. 3a).

To examine whether reduced expression of DNMT genes affects siRNA-mediated DNA methylation, we introduced a combination of E-cadherin-siRNAs (sites 1–10) into MCF-7 cells transfected with *DNMT1*-, *DNMT2*- or *DNMT3B*-siRNAs. As shown in Fig. 3b, the extent of DNA methylation induced by E-cadherin-siRNAs in MCF-7 cells transfected with *DNMT1*- and *DNMT3B*-siRNA was significantly lower than in wild-type MCF-7 cells or in MCF-7 cells transfected with mutant *DNMT*-siRNA. By contrast, disruption of



**Figure 2** Effects of E-cadherin-siRNAs targeted to the E-cadherin promoter on E-cadherin mRNA expression. **a**, Levels of E-cadherin mRNA expression in cells transfected with E-cadherin-siRNAs. Total RNA from each line was analysed by northern blotting analysis. Actin mRNA was used as the endogenous control. Values are means  $\pm$  s.d. **b**, Relative levels of E-cadherin mRNA expression in cells transfected with E-cadherin-siRNAs and

grown with and without 5-aza. **c**, Levels of E-cadherin expression in cells transfected with E-cadherin-siRNAs. E-cadherin in each line was detected by western blotting with specific antibodies. Relative levels of E-cadherin were analysed densitometrically. Values are means  $\pm$  s.d.

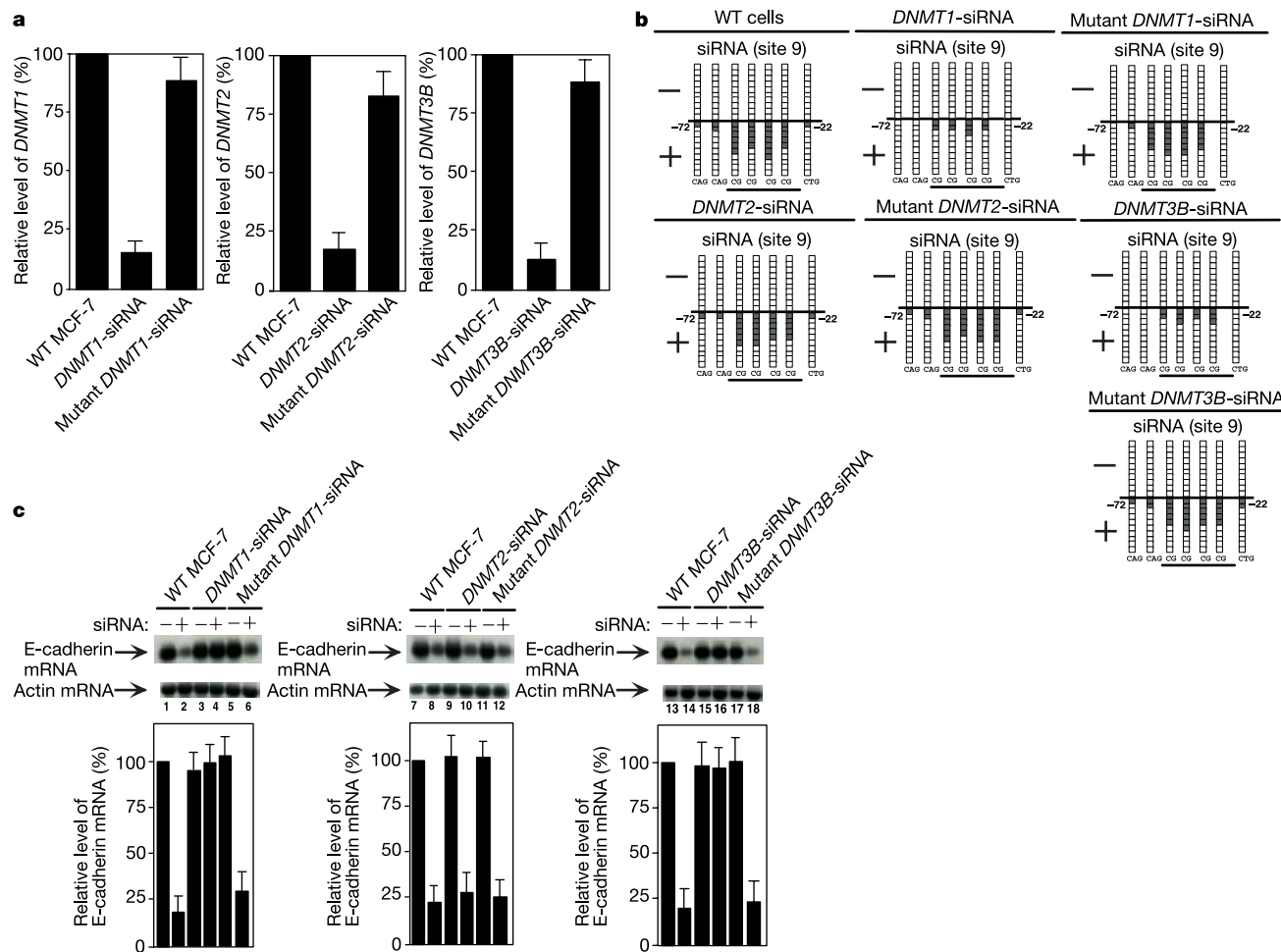
*DNMT2* expression did not significantly affect siRNA-mediated DNA methylation. Thus, *DNMT1* and *DNMT3B*, but not *DNMT2*, seem to be necessary for siRNA-mediated DNA methylation in human cells.

We next examined the effects of E-cadherin-siRNAs on the expression of E-cadherin in MCF-7 cells transfected with *DNMT*-siRNA. As shown in Fig. 3c, siRNAs targeted to the E-cadherin promoter did not alter E-cadherin gene expression in MCF-7 cells that contained siRNAs targeted to *DNMT1* or *DNMT3B* mRNA (lanes 4 and 16). By contrast, E-cadherin-siRNAs suppressed E-cadherin gene expression in MCF-7 cells that contained *DNMT2*-siRNA or mutant *DNMT*-siRNAs (lanes 6, 10, 12 and 18), demonstrating that *DNMT1* and *DNMT3B* are necessary for the transcriptional gene silencing that results from the induction of DNA methylation by E-cadherin-siRNAs.

To examine the possibility for gene therapy via the control of DNA methylation by siRNAs, we constructed expression vectors for short hairpin RNAs (shRNAs) targeted to an *erbB2* promoter. The *erbB2* gene is overexpressed and unmethylated in several lines of tumour cells, such as MCF-7 cells, whereas it can be silenced by methylation of the promoter in several lines of normal cells<sup>26</sup>. To express shRNAs in cells, we used the well-characterized tRNA<sup>Val</sup> promoter system<sup>16</sup>. We have demonstrated previously that tRNA<sup>Val</sup>-driven shRNA induces siRNA-mediated gene silencing in

human cells<sup>16</sup>. We selected five CpG islands (sites 1–5) within the *erbB2* promoter as targets of shRNAs. Then, we introduced these shRNA expression plasmids transiently into MCF-7 cells. We confirmed by means of northern blotting that appropriate processing and production of siRNAs had occurred in cells that expressed tRNA-shRNAs (Fig. 4b, lanes 1–5), and then we examined levels of DNA methylation within the *erbB2* promoter. We isolated genomic DNA from each cell line and performed bisulphite sequencing. As shown in Fig. 4c, we detected DNA methylation within the *erbB2* promoter in all lines of MCF-7 cells that expressed a tRNA-shRNA or a combination of all tRNA-shRNAs (sites 1–5). However, a mutant tRNA-shRNA (site 1) with nine point mutations in both the sense and antisense strand did not induce DNA methylation within the *erbB2* promoter. In addition, tRNA-shRNA did not induce DNA methylation within the non-targeted E-cadherin promoter (Supplementary Fig. S3). We obtained similar results with U6 promoter-driven shRNAs (data not shown), demonstrating that induction of sequence-specific DNA methylation by vector-based siRNAs in human cells is a general phenomenon.

To determine whether the vector-based shRNAs had suppressed the expression of *erbB2*, we examined levels of *erbB2* mRNA expression by means of northern blotting. As shown in Fig. 4d, the expression level of *erbB2* mRNA in cells that expressed each respective tRNA-shRNA was lower than in wild-type MCF-7 cells.



**Figure 3** Effects of E-cadherin-siRNAs targeted to the E-cadherin promoter in *DNMT1*-, *DNMT2*- and *DNMT3B*-knockdown cells. **a**, Amount of DNMTs in MCF-7 cells in the presence or absence of *DNMT*-siRNAs. The amount of *DNMT1*, *DNMT2* and *DNMT3B* was detected by western blotting. Values are means  $\pm$  s.d. **b**, Level of DNA methylation of the E-cadherin promoter in MCF-7 cells in the presence or absence of E-cadherin-siRNA and

various *DNMT*-siRNAs. DNA methylation was detected by bisulphite sequencing. **c**, E-cadherin mRNA expression levels in MCF-7 cells in the presence or absence of E-cadherin-siRNA (sites 1–10) and various *DNMT*-siRNAs. Total RNA from each line was analysed by northern blotting with specific probes. Values are means  $\pm$  s.d.

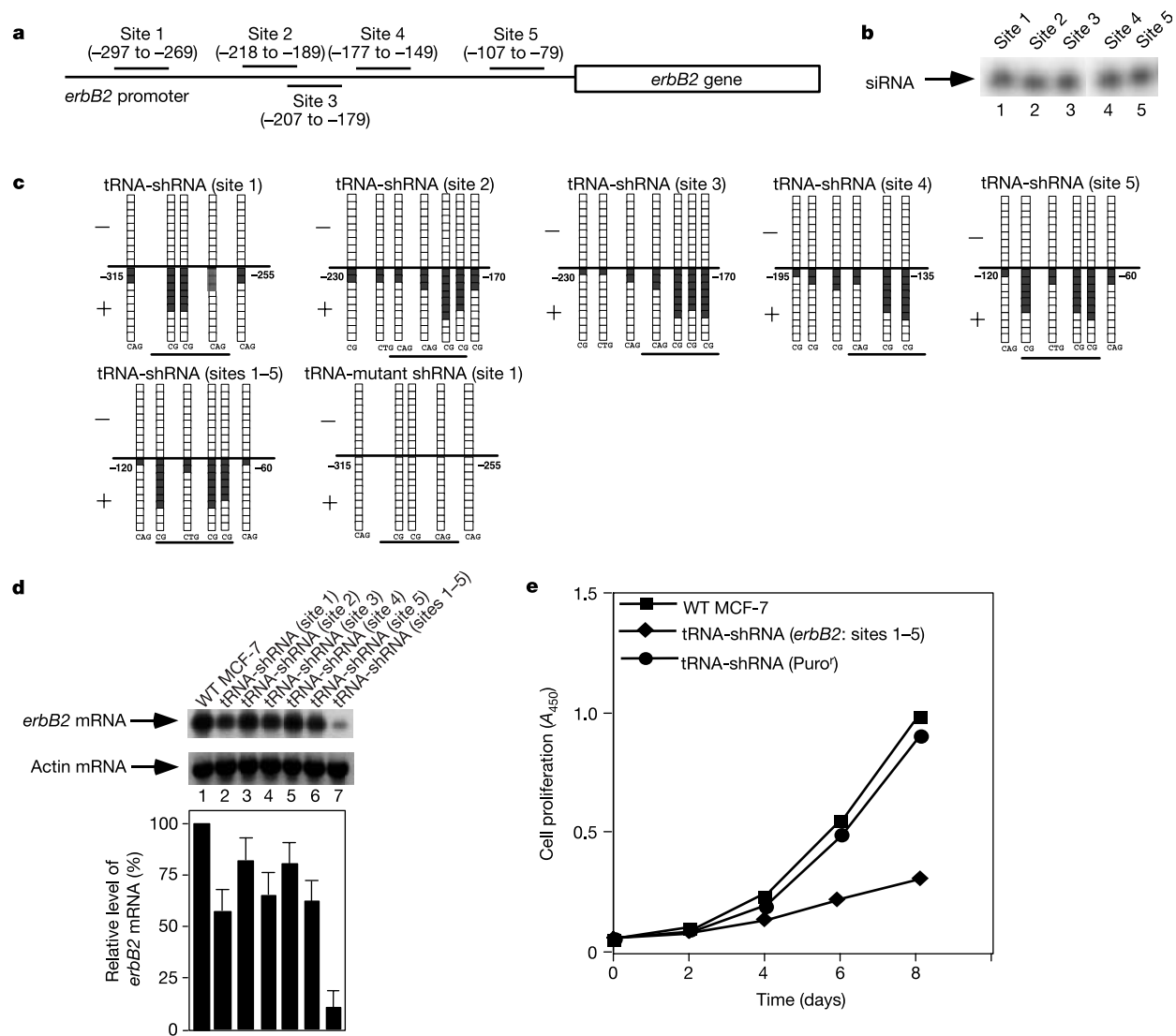
As observed previously (Fig. 2a), the level of *erbB2* mRNA expression in cells that expressed all tRNA-shRNAs (sites 1–5) together was significantly lower than in wild-type MCF-7 cells and in cells that expressed individual tRNA-shRNAs, demonstrating the additivity of suppression by vector-based shRNAs at the transcriptional level. Transcriptional regulation through the induction of methylation within a promoter by siRNA in mammalian cells seems to be a general phenomenon, as demonstrated by the examples in this study.

To examine the phenotype of cells that expressed tRNA-shRNAs, we analysed the proliferation rates of various cell lines. MCF-7 cells expressing all tRNA-shRNAs (sites 1–5) proliferated significantly more slowly than wild-type MCF-7 cells and cells that expressed tRNA-shRNAs targeted to a puromycin-resistance gene, which we chose as a nonspecific control (Fig. 4e). The reduced proliferation rate of MCF-7 cells that expressed all tRNA-shRNAs (sites 1–5) was correlated with a reduction in the level of *erbB2* mRNA expression in these cells, suggesting that the tRNA-shRNAs targeted to the *erbB2* promoter might have potential utility as therapeutic agents.

The DNA methylation of promoters has an important role in the genesis and development of tumours by regulating the expression of specific genes<sup>27,28</sup>. Many proteins involved in DNA methylation, such as MeCP2, MBD and DNMTs, have been well characterized in human cells. However, it remains unclear how and under what circumstances they are guided to and methylate specific CpG target sites of cognate genes during the genesis or development of tumours.

In plants, long and short dsRNAs can induce sequence-specific DNA methylation, known as RNA-directed DNA methylation<sup>3–8</sup>. In addition, transgenes can also induce sequence-specific DNA methylation<sup>29,30</sup>. These phenomena might reflect the role of these systems as a cellular defence against RNA and DNA viruses. However, it remains to be determined whether similar RNA-directed DNA methylation and transgene-mediated defence systems exist in human cells.

We have demonstrated that synthetic and vector-based siRNAs can induce sequence-specific and *DNMT1/DNMT3B*-dependent RNA-mediated DNA methylation in human cells. Our siRNAs



**Figure 4** Induction of DNA methylation of the *erbB2* promoter by tRNA-shRNAs. **a**, Five targets for tRNA-shRNAs were selected within the *erbB2* promoter. **b**, Detection, via Northern blots, of the expression of tRNA-shRNAs targeted to the *erbB2* promoter. All anticipated siRNAs were detected in cells that expressed tRNA-shRNAs. **c**, Induction by shRNAs of DNA methylation of the *erbB2* promoter, as detected by bisulphite sequencing.

**d**, Relative levels of *erbB2* mRNA in cells that expressed tRNA-shRNAs targeted to the *erbB2* promoter. Total RNA from each line was analysed by Northern blotting with specific probes. Values are means  $\pm$  s.d. **e**, Proliferation rates of cells expressing tRNA-shRNAs. tRNA-shRNA (puro'), targeted to a puromycin-resistance gene, was used as a control shRNA.



induced sequence-specific gene silencing at the transcriptional level. The ways in which siRNAs are guided to and gain access to genomic DNA remain unknown. Synthetic and tRNA vector-based siRNAs are localized predominantly in the cytoplasm, where siRNA-mediated degradation of mRNA also occurs. A small fraction of siRNA-protein complexes might be transported to the nucleus. Alternatively, siRNAs might gain access to genomic DNA during cell division, when the nuclear membrane disappears.

Further investigation of correlations between siRNA-induced DNA methylation and tumour genesis, anti-viral defence and epigenesis in development should provide an insight into small-RNA-induced gene silencing and development. Moreover, it is now possible to control the expression levels of specific genes in mammalian cells using siRNAs not only to disrupt cognate mRNAs but also to interfere with transcription, as demonstrated here. However, it remains to be examined precisely which promoters can be specifically methylated by respective siRNAs. Exploitation of shRNA expression vectors targeted to cognate promoters might even have potential utility in a clinical setting. □

## Methods

### Preparation of siRNAs

Synthetic siRNAs directed against the E-cadherin promoter (E-cadherin-siRNA) and against DNMT mRNAs (DNMT-siRNAs) were synthesized with a DNA/RNA synthesizer (model 394; PE Applied Biosystems). Sequences of E-cadherin-siRNAs are described in the Supplementary Information. For generation of siRNAs, all synthetic RNAs were annealed by the standard method<sup>13</sup>. We introduced siRNAs into MCF-7 cells using Oligofectamine (Invitrogen) in accordance with the manufacturer's protocol.

### Construction of tRNA-shRNA expression plasmids

To construct vectors for expression of tRNA-shRNA targeted to the *erbB2* promoter, we used the pPUR-tRNA plasmid, which includes a chemically synthesized promoter for a human gene for tRNA<sup>Val</sup> between the *EcoRI* and *BamHI* sites of pPUR (Clontech)<sup>16</sup>. Sequences (sites 1–5) of shRNAs targeted to the *erbB2* promoter are described in Supplementary Information. Chemically synthesized oligonucleotides encoding *erbB2* promoter-directed shRNAs that included a loop motif were amplified as double-stranded sequences by PCR. After digestion with *SacI* and *KpnI*, the fragments were cloned downstream of the promoter of the tRNA<sup>Val</sup> gene in pPUR-tRNA.

### Bisulphite sequencing of E-cadherin and *erbB2* promoters

We extracted genomic DNA from cells by standard methods using proteinase K, phenol and chloroform. We performed bisulphite modifications using a CpGenome DNA modification kit (Intergen) following the manufacturer's instructions. We amplified the bisulphite-modified E-cadherin and *erbB2* promoters using specific primers as follows: for the E-cadherin promoter, forward primer 5'-TCTAGAAAAATTTTAAAAA-3' and reverse primer 5'-CAGCGCCGAGAGGCTGCGGCT-3'; for the *erbB2* promoter, forward primer 5'-CCTGGAAGCCCAAGGTAAAC-3' and reverse primer 5'-TTTCTCCGGTCCCAATGGAGG-3'. The amplified DNAs were subcloned into the TA-cloning vector. Then we picked ten independent colonies in each case, determined the sequence of the promoter in each plasmid and examined the extent of methylation by determining the number and position of methylated cytosine residues.

### Culture and transfection of cells

Human MCF-7 cells were cultured in minimum essential medium (MEM) supplemented with 10% FCS, 1% NEAA and 1 mM Na-Pyr. Human normal mammary epithelial cells (CAMBREX) were grown with mammary epithelial cell medium (CAMBREX). Transfection with pPUR-tRNA-shRNA was performed with the Effectene reagent (Qiagen) according to the manufacturer's protocol. For transfection of MCF-7 cells with various siRNAs, we used Oligofectamine (Invitrogen) according to the manufacturer's protocol.

### 5-aza-2'-deoxycytidine treatment

For inhibition of DNA methylation by 5-aza-2'-deoxycytidine (5-aza), cells grown to 30–40% confluence were transfected with E-cadherin-siRNAs and then were incubated for 96 h in culture medium that contained a final concentration of 1  $\mu$ M 5-aza (Sigma).

### Northern blotting analysis

Total RNA was purified from MCF-7 cells that expressed tRNA-shRNAs targeted to the *erbB2* promoter with ISOGEN reagent (Wako). Thirty micrograms of total RNA per lane were loaded on a 15% polyacrylamide gel. After electrophoresis, bands of RNA were transferred to a Hybond-N nylon membrane (Amersham). The RNA on the membrane was allowed to hybridize to <sup>32</sup>P-labelled probes that were complementary to the sequences of the tRNA-shRNAs. Sequences of synthetic probes were as follows: site 1, 5'-CUCUGCCCCUCCCCGAGUCCGGAUAA-3'; site 2, 5'-UCCUAGCGCGGGAAGCU GGGUUGCCUGCA-3'; site 3, 5'-GGUGCGUCCCUAGCGCGGGAAGCUGG-3'; site 4, 5'-GGUGCGUCCCUAGCGCGGGAAGCUGG-3'; and site 5, 5'-GAGCAA GCGCGUCCAGCUGCCCCUCC-3'. For detection of the expression level of

E-cadherin, *erbB2* and actin mRNAs, we used cDNA probes of E-cadherin, *erbB2* and actin, respectively.

### Western blotting analysis

MCF-7 cells transfected with or without various E-cadherin-siRNAs targeted to the E-cadherin promoter and various tRNA-shRNAs targeted to the *erbB2* promoter were collected. Total protein (20  $\mu$ g) was fractionated by SDS-polyacrylamide gel electrophoresis (10% polyacrylamide) and bands of protein were transferred to a polyvinylidene difluoride membrane (Funakoshi) by electro-blotting. Immunocomplexes were visualized with an ECL kit (Amersham) after reactions with monoclonal antibodies against E-cadherin (Transduction laboratories), ErbB2 (Oncogene), actin (Chemicon), DNMT1 (Imgenex) and DNMT3B (Imgenex), or with polyclonal antibodies against DNMT2 (Imgenex). Amounts of E-cadherin and ErbB2 were normalized by reference amounts of actin.

### Determination of cell proliferation rates

Cell proliferation rates were determined with a Cell Proliferation Kit II (Roche) according to the manufacturer's instructions<sup>16</sup>.

Received 18 March; accepted 26 July 2004; doi:10.1038/nature02889.

Published online 15 August 2004.

1. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
2. Hutvagner, G. & Zamore, P. D. RNAi: nature abhors a double-strand. *Curr. Opin. Genet. Dev.* **12**, 225–232 (2002).
3. Pelissier, T. & Wassenegger, M. A DNA target of 30 bp is sufficient for RNA-directed DNA methylation. *RNA* **6**, 55–65 (2000).
4. Mette, M. F., Aufsatz, W., van der Winden, J., Matzke, M. A. & Matzke, A. J. Transcriptional silencing and promoter methylation triggered by double-stranded RNA. *EMBO J.* **19**, 5194–5201 (2000).
5. Jones, L., Ratcliff, F. & Baulcombe, D. C. RNA-directed transcriptional gene silencing in plants can be inherited independently of the RNA trigger and requires Met1 for maintenance. *Curr. Biol.* **11**, 747–757 (2001).
6. Hamilton, A., Voinnet, O., Chappell, L. & Baulcombe, D. Two classes of short interfering RNA in RNA silencing. *EMBO J.* **21**, 4671–4769 (2002).
7. Aufsatz, W., Mette, M. F., Van Der Winden, J., Matzke, M. & Matzke, A. J. HDA6, a putative histone deacetylase needed to enhance DNA methylation induced by double-stranded RNA. *EMBO J.* **21**, 6832–6841 (2002).
8. Zilberman, D., Cao, X. & Jacobsen, S. E. ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science* **299**, 716–719 (2003).
9. Bernstein, E., Caudy, A. A., Hammond, S. M. & Hannon, G. J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363–366 (2001).
10. Elbashir, S. M., Lendeckel, W. & Tuschl, T. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.* **15**, 188–200 (2001).
11. Zamore, P., Tuschl, T., Sharp, P. & Bartel, D. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21- to 23-nucleotide intervals. *Cell* **101**, 25–33 (2000).
12. Hammond, S. M., Bernstein, E., Beach, D. & Hannon, G. J. An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature* **404**, 293–296 (2000).
13. Elbashir, S. M. *et al.* Duplexes of 21-nucleotide RNAs mediate RNA interference in mammalian cell culture. *Nature* **411**, 494–498 (2001).
14. Hutvagner, G. & Zamore, P. D. A microRNA in a multiple-turnover RNAi enzyme complex. *Science* **297**, 2056–2060 (2002).
15. Zeng, Y. & Cullen, B. R. RNA interference in human cells is restricted to the cytoplasm. *RNA* **8**, 855–860 (2002).
16. Kawasaki, H. & Taira, K. Short hairpin type of dsRNAs that are controlled by tRNA<sup>Val</sup> promoter significantly induce RNAi-mediated gene silencing in the cytoplasm of human cells. *Nucleic Acids Res.* **31**, 700–707 (2003).
17. Herman, J. G., Graff, J. R., Myohanen, S., Nelkin, B. D. & Baylin, S. B. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc. Natl Acad. Sci. USA* **93**, 9821–9826 (1996).
18. Graff, J. R., Herman, J. G., Myohanen, S., Baylin, S. B. & Vertino, P. M. Mapping patterns of CpG island methylation in normal and neoplastic cells implicates both upstream and downstream regions in *de novo* methylation. *J. Biol. Chem.* **272**, 22322–22329 (1997).
19. Corn, P. G. *et al.* E-cadherin expression is silenced by 5' CpG island methylation in acute leukemia. *Clin. Cancer Res.* **6**, 4243–4248 (2000).
20. Volpe, T. A. *et al.* Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**, 1833–1837 (2002).
21. Pal-Bhadra, M. *et al.* Heterochromatic silencing and HP1 localization in *Drosophila* are dependent on the RNAi machinery. *Science* **303**, 669–672 (2004).
22. Maison, C. *et al.* Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. *Nature Genet.* **30**, 329–334 (2002).
23. Bestor, T. H. The DNA methyltransferases of mammals. *Hum. Mol. Genet.* **9**, 2395–2402 (2000).
24. Rhee, I. *et al.* DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. *Nature* **416**, 552–556 (2002).
25. Robert, M. F. *et al.* DNMT1 is required to maintain CpG methylation and aberrant gene silencing in human cancer cells. *Nature Genet.* **33**, 61–65 (2003).
26. Hattori, M., Sakamoto, H., Satoh, K. & Yamamoto, T. DNA demethylase is expressed in ovarian cancers and the expression correlates with demethylation of CpG sites in the promoter region of *c-erbB-2* and *survivin* genes. *Cancer Lett.* **169**, 155–164 (2001).
27. Li, E. Chromatin modification and epigenetic reprogramming in mammalian development. *Nature Rev. Genet.* **3**, 662–673 (2002).
28. Esteller, M. CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future. *Oncogene* **21**, 5427–5440 (2002).
29. Morel, J. B., Mourrain, P., Beclin, C. & Vaucheret, H. DNA methylation and chromatin structure affect transcriptional and post-transcriptional transgene silencing in *Arabidopsis*. *Curr. Biol.* **10**, 1591–1594 (2000).

30. Beclin, C., Boutet, S., Waterhouse, P. & Vaucheret, H. A branched pathway for transgene-induced RNA silencing in plants. *Curr. Biol.* **12**, 684–688 (2002).

Supplementary Information accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This research was supported by grants from AIST and by a Grant-in-Aid for Scientific Research and for the 21st Century COE programmes, Center for Integrated Brain Medical Science and Human-Friendly Materials based on Chemistry, from the Ministry of Education, Culture, Sports, Science and Culture (MEXT) of Japan.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to K.T. (taira@chembio.t.u-tokyo.ac.jp) or H.K. (kawasaki@chembio.t.u-tokyo.ac.jp).

## Error-prone replication of oxidatively damaged DNA by a high-fidelity DNA polymerase

Gerald W. Hsu<sup>1</sup>, Matthias Ober<sup>2</sup>, Thomas Carell<sup>2</sup> & Lorena S. Beese<sup>1</sup>

<sup>1</sup>Department of Biochemistry, Duke University Medical Center, Durham, North Carolina 27710, USA

<sup>2</sup>Department of Chemistry and Biochemistry, Ludwig Maximilians University Munich, Butenandtstrasse 5-13, D 81377 Munich, Germany

Aerobic respiration generates reactive oxygen species that can damage guanine residues and lead to the production of 8-oxoguanine (8oxoG), the major mutagenic oxidative lesion in the genome<sup>1</sup>. Oxidative damage is implicated in ageing<sup>2</sup> and cancer, and its prevalence presents a constant challenge to DNA polymerases that ensure accurate transmission of genomic information. When these polymerases encounter 8oxoG, they frequently catalyse misincorporation of adenine in preference to accurate incorporation of cytosine<sup>3</sup>. This results in the propagation of G to T transversions, which are commonly observed somatic mutations associated with human cancers<sup>4,5</sup>. Here, we present sequential snapshots of a high-fidelity DNA polymerase during both accurate and mutagenic replication of 8oxoG. Comparison of these crystal structures reveals that 8oxoG induces an inversion of the mismatch recognition mechanisms that normally proofread DNA, such that the 8oxoG:adenine mismatch mimics a cognate base pair whereas the 8oxoG:cytosine base pair behaves as a mismatch. These studies reveal a fundamental mechanism of error-prone replication and show how 8oxoG, and DNA lesions in general, can form mismatches that evade polymerase error-detection mechanisms, potentially leading to the stable incorporation of lethal mutations.

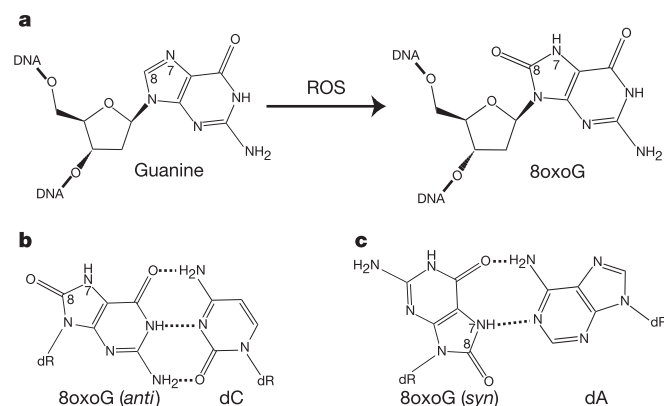
Many DNA damage lesions stall or block DNA replication; however, 8oxoG is bypassed efficiently and inaccurately by high-fidelity polymerases<sup>3,6–8</sup>. 8oxoG retains the ability to engage in correct Watson–Crick base pairs with C, but oxidation of G (at C8) converts a hydrogen bond acceptor (N7) to a hydrogen bond donor, allowing a stable Hoogsteen base pair to form between 8oxoG and A, which is not possible in undamaged DNA (Fig. 1). In the absence of accessory factors such as proliferating cell nuclear antigen (PCNA), preferential mutagenic translesion replication of 8oxoG by the major replicative DNA polymerases  $\alpha$  and  $\delta$  *in vitro* indicates that 8oxoG is not recognized as a DNA lesion. Instead, the A·8oxoG mismatch evades the intrinsic mechanisms that cause DNA polymerases to achieve high-fidelity replication of undamaged DNA.

Insights into the mechanisms of faithful replication by high-

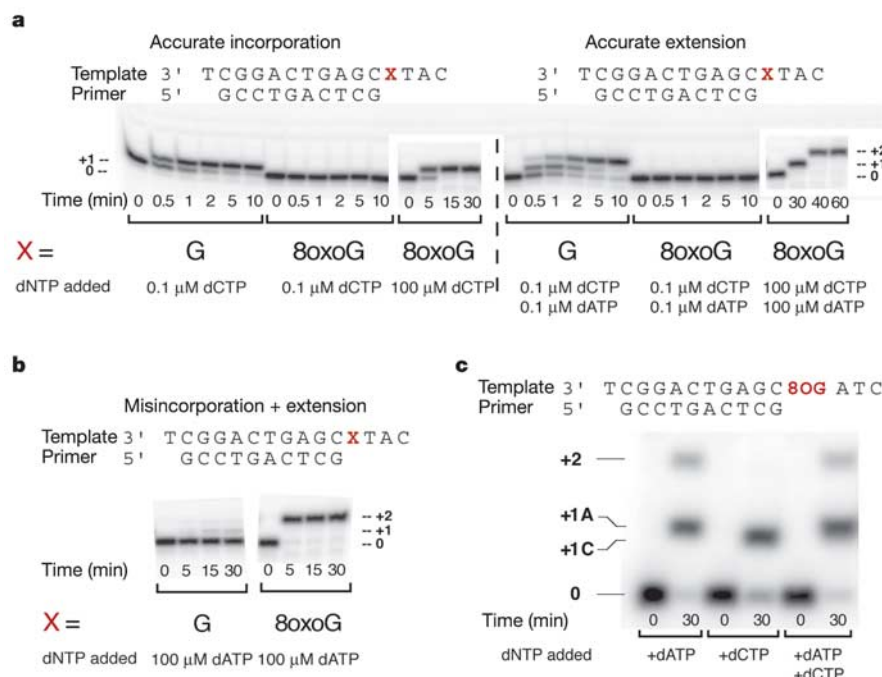
fidelity polymerases<sup>9</sup> and mismatch recognition<sup>10</sup> have been provided by capturing structures of the replication cycle of the DNA polymerase I fragment from a thermostable strain of *Bacillus stearothermophilus* (BF) in a polymerase crystal that retains the ability to replicate DNA<sup>11</sup>. Throughout the replicative cycle of BF, there are many points at which recognition of DNA mismatches or lesions can occur. First, during the transition of the template strand from a pre-insertion site that sequesters the template base before nucleotide incorporation to an insertion site where the template base interacts with the incoming nucleotide, the conformations of the template base and the incoming nucleotide are both tightly regulated to ensure that hydrogen bonding interactions are limited to the Watson–Crick faces of the nascent base pair. Second, before covalent incorporation of the paired dNTP at the insertion site, the polymerase selects for base pairs that exhibit the shape and geometry of cognate Watson–Crick base pairs in preference to ones that do not<sup>12,13</sup>. Third, after covalent incorporation, the new base pair moves to a post-insertion site where residues Arg 615 and Gln 797 form hydrogen bonds to the DNA minor groove of correctly formed base pairs. At this step, DNA mismatches induce distortions within the polymerase active site that cause the polymerase to stall and dissociate<sup>10</sup>. Last, covalently incorporated mismatches or lesions continue to impede replication from up to four base pairs away<sup>14</sup>, inducing distortions to the active site as they translocate along the surface of the polymerase<sup>10</sup>. In this way, a molecular ‘memory’ of the mismatch is retained by the polymerase.

To investigate whether BF behaves analogously to replicative polymerases  $\alpha$  and  $\delta$  with respect to replication of 8oxoG, primer extension assays were performed (Fig. 2) and steady-state kinetic parameters were determined for incorporation of dCTP or dATP opposite 8oxoG (Table 1). Comparison of the specificity constants ( $k_{\text{cat}}/K_m$ ; where  $k_{\text{cat}}$  is the turnover number and  $K_m$  is the Michaelis constant) for incorporation of these nucleotides opposite 8oxoG indicates that misincorporation of dATP is ninefold more efficient than dCTP incorporation, whereas in undamaged DNA, accurate dCTP incorporation opposite an unmodified guanine is 10<sup>6</sup>-fold more efficient than misincorporation of dATP. After incorporation of either dATP or dCTP, extension past the newly formed 8oxoG base pair is readily observed (Fig. 2a, b) and proceeds to the end of the template in the presence of a full complement of dNTPs (Supplementary Fig. 1). These results indicate that BF, as with replicative polymerases  $\alpha$  and  $\delta$ , preferentially misincorporates dATP opposite 8oxoG. BF can therefore serve as a model system to address mutagenic 8oxoG replication.

To determine the effects of 8oxoG on the structural mechanisms



**Figure 1** Modes of base pairing for 8oxoG. **a**, Oxidation of guanine at C8 by reactive oxygen species (ROS). **b**, 8oxoG in a Watson–Crick base pair with dC. Dashed lines indicate potential hydrogen bonds. **c**, 8oxoG (*syn*) in a Hoogsteen base pair with dA (*anti*).



**Figure 2** Translesion replication of 8oxoG by BF in solution. **a**, **b**, Accurate (**a**) and mutagenic (**b**) replication of 8oxoG in solution. Primers (10 nucleotides) annealed to templates containing unmodified guanine or 8oxoG at position X were extended by one or two nucleotides. To observe replication of 8oxoG, dNTP concentrations were raised to 100  $\mu$ M. Extension of C-8oxoG was performed by allowing incorporation of dCTP to occur

to completion (30 min) before adding dATP so that extended products exclusively represent extension of C-8oxoG. **c**, In the presence of equimolar amounts of dATP and dCTP (last lane), BF preferentially misincorporates dATP. A quantitative analysis is shown in Table 1.

of mismatch recognition, we obtained sequential structural views of BF as it accurately (Fig. 3; see also Supplementary Fig. 2) and mutagenically (Fig. 4; see also Supplementary Figs 2 and 3) replicated 8oxoG (Table 2). These structures show 8oxoG at the pre-insertion site before nucleotide incorporation, at the post-insertion site after incorporation of dCTP or dATP, and in the DNA duplex binding region after extension of 8oxoG-containing base pairs. The polymerase structure with 8oxoG at the pre-insertion site (Fig. 3b) is similar to that with all other template bases at this position. At the post-insertion site, 8oxoG forms a Watson–Crick base pair with cytosine (Fig. 3c) but a Hoogsteen base pair with adenine (Fig. 4a). At this position, the C-8oxoG (primer:template) base pair disrupts the polymerase active site, but the A-8oxoG base pair does not. However, when either the C-8oxoG or A-8oxoG base pairs occupy the DNA duplex binding region after extension (Figs 3d and 4b), neither distorts the active site.

At the pre-insertion site, 8oxoG resembles a cognate template base and adopts a *syn* conformation (Fig. 3b). Transition of the template base from the pre-insertion site is typically associated with a change to an *anti* conformation at the post-insertion site. This transition remains intact upon incorporation of dCTP but not dATP. By adopting an *anti* conformation to pair opposite cytosine (Fig. 3c), 8oxoG induces template distortion to avoid a potential steric clash between the C8 carbonyl oxygen and the O4' of the sugar moiety. The relative orientation of the sugar and base is altered, thereby inducing distortion to the template 5' and 3' of the lesion. This template distortion is transmitted to the polymerase active site, resulting in active site distortions that resemble those induced by mismatches in undamaged DNA<sup>10</sup>. Template distortions induced by 8oxoG have previously been reported in ternary complex structures of RB69 (ref. 8) and DNA polymerase  $\beta$ <sup>15</sup> bound to an 8oxoG-containing template and an incoming dCTP, indicating that tem-

plate distortions are initiated before catalysis of the phosphodiester bond and are an intrinsic property of the lesion due to internal steric strain. Here we observe that these distortions affect not only the template but the polymerase as well. The 8oxoG-containing template lifts away from the surface of the polymerase, disrupting the interaction with the minor-groove reading residue Gln 797, and the polymerase O and O1 helices adopt a distorted conformation that prevents the next template base from occupying the pre-insertion site. Similar template and polymerase distortions are observed when DNA mismatches occupy the post-insertion site<sup>10</sup>, and account for the inhibitory effect of mismatches on further DNA synthesis<sup>16–18</sup>.

Template and polymerase distortions are not observed when 8oxoG adopts a *syn* conformation that enables it to form a Hoogsteen base pair with adenine. In this arrangement, the active site is not disrupted (Fig. 4a). By adopting a *syn* conformation, 8oxoG eliminates internal steric strain, which would cause template distortion. In the Hoogsteen base-pairing arrangement, the surface of the A-8oxoG mismatch resembles a normal Watson–Crick base pair in shape and geometry, and occupies the post-insertion site without distorting the protein, template, or primer. Furthermore, hydrogen bonds between polymerase residues Arg615 and Gln 797 and the

**Table 1** Steady-state incorporation kinetics

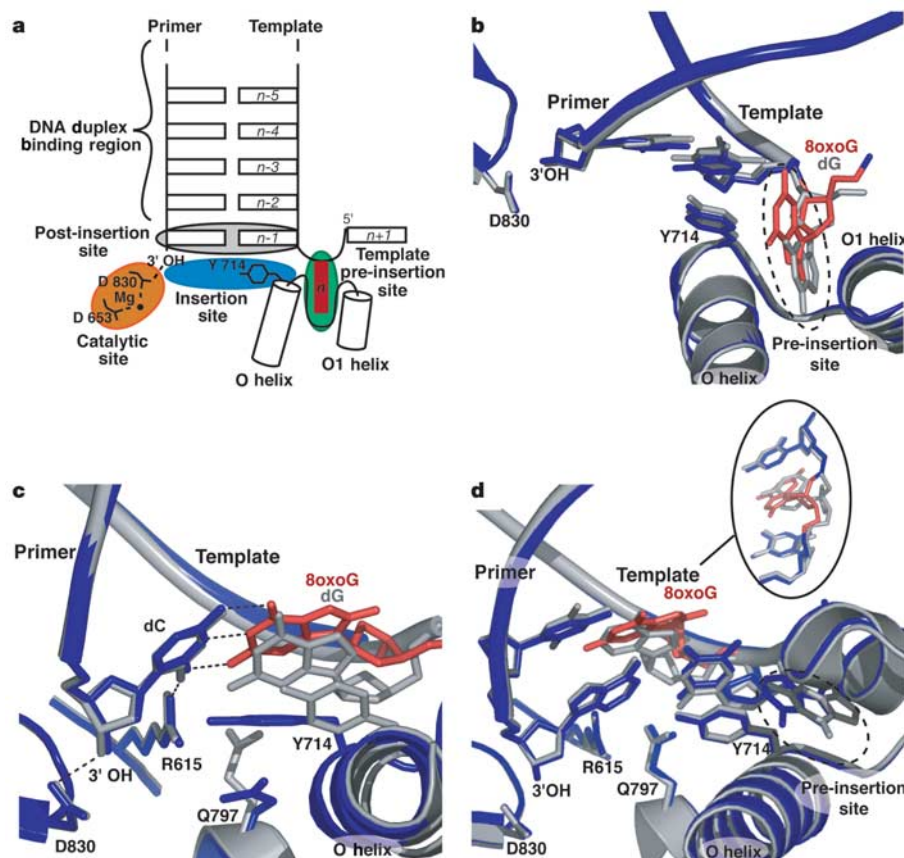
| Template | dNTP | $K_m$<br>( $\mu$ M) | $k_{cat}$<br>( $\text{min}^{-1}$ ) | $k_{cat}/K_m$<br>( $\mu\text{M}^{-1}\text{min}^{-1}$ ) | dCTP:dATP*        |
|----------|------|---------------------|------------------------------------|--------------------------------------------------------|-------------------|
| 8oxoG    | dCTP | 100.7               | 13                                 | 0.13                                                   | 0.11              |
|          | dATP | 46.9                | 54                                 | 1.15                                                   |                   |
| dG       | dCTP | 1.5                 | 750                                | 500                                                    | $7.9 \times 10^6$ |
|          | dATP | 111.1               | 0.007                              | $6.3 \times 10^{-5}$                                   |                   |

\*Values of  $k_{cat}/K_m$  ratio.



A·8oxoG base pair are retained, as the minor groove of A·8oxoG is virtually identical in geometry to the minor groove of a cognate A·T base pair. Therefore, the A·8oxoG mismatch evades mismatch detection. Although Hoogsteen base pairing has been observed to facilitate accurate replication by error prone polymerases<sup>19,20</sup>, and

has been predicted to allow 8oxoG to form a base pair with A<sup>8,15,21,22</sup>, the observation of a template base in a *syn* conformation at the high-fidelity polymerase active site has not been reported previously. Formation of the 8oxoG-induced Hoogsteen base pair is particularly insidious in this context because it promotes the formation of a



**Figure 3** Accurate translesion replication of 8oxoG by BF in crystals. **a**, Schematic of BF active site. During replication, the template base (*n*, red) moves from the pre-insertion site to the post-insertion site to the DNA duplex binding region. **b–d**, Structures of accurate 8oxoG replication (blue) are superimposed with structures of unmodified guanine

replication (grey). The 8oxoG template base (red) is shown at the pre-insertion site (**b**) before nucleotide incorporation, the post-insertion site (**c**) after dCTP incorporation, and the DNA duplex binding region (**d**) after extension of C·8oxoG.

**Table 2** Summary of crystallographic data

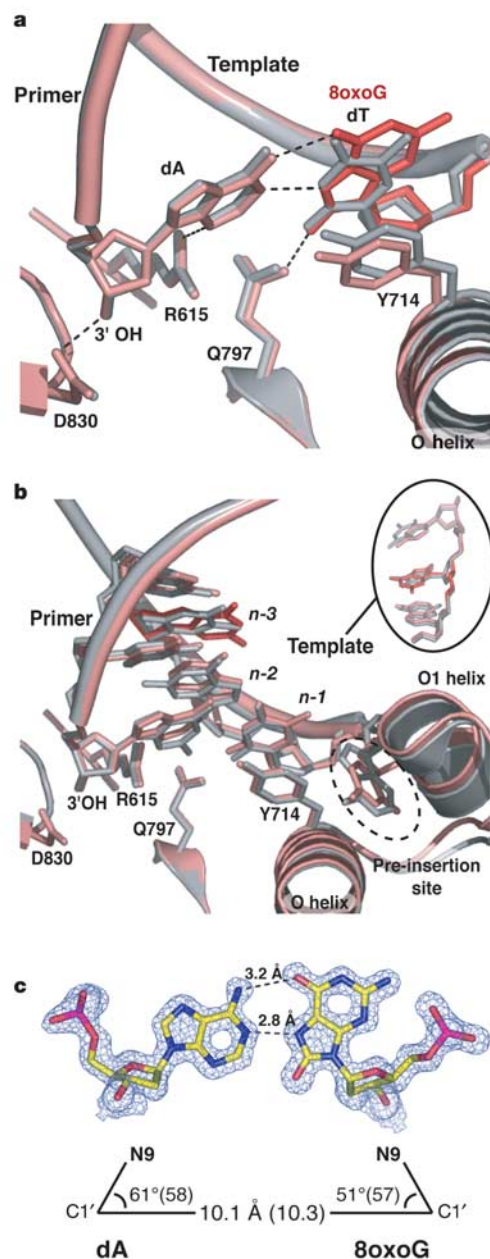
| Crystal                                      | 8oxo        | C-8oxo      | C-8oxo      | A-8oxo      | A-8oxo      |
|----------------------------------------------|-------------|-------------|-------------|-------------|-------------|
| Position of lesion                           | Pre-IS      | Post-IS     | <i>n</i> -2 | Post-IS     | <i>n</i> -3 |
| Wavelength (Å)                               | 1.542†      | 1.542†      | 1.542†      | 1.542†      | 1.000‡      |
| Resolution range (Å)                         | 50–2.0      | 50–2.0      | 50–2.1      | 50–2.15     | 50–1.6      |
| Outer shell (Å)                              | 2.07–2.0    | 2.07–2.0    | 2.18–2.1    | 2.23–2.15   | 1.66–1.6    |
| <i>R</i> <sub>sym</sub> (%)                  | 5.0 (12.2)  | 9.1 (30.2)  | 7.5 (24.4)  | 6.5 (22.4)  | 7.3 (19.0)  |
| Unique reflections                           | 58,360      | 59,296      | 51,291      | 48,131      | 114,192     |
| Total reflections                            | 496,318     | 838,647     | 668,787     | 623,476     | 1,859,662   |
| Mean <i>I</i> / <i>σ</i> <sub><i>I</i></sub> | 20.3 (6.4)  | 11.0 (2.6)  | 10.1 (2.6)  | 12.3 (3.0)  | 16.3 (4.4)  |
| Completeness (%)                             | 92.1 (81.6) | 92.9 (86.4) | 89.5 (82.9) | 90.9 (81.4) | 93.8 (75.5) |
| <i>R</i> <sub>cryst</sub> (%)                | 19.4        | 21.6        | 20.5        | 20.5        | 21.3        |
| <i>R</i> <sub>free</sub> (%)                 | 22.7        | 24.7        | 23.9        | 24.7        | 22.7        |
| r.m.s. deviation bond length (Å)             | 0.005       | 0.005       | 0.006       | 0.006       | 0.005       |
| r.m.s. deviation bond angle (deg)            | 1.2         | 1.2         | 1.2         | 1.2         | 1.2         |
| Average <i>B</i> -value (Å <sup>2</sup> )    | 24.1        | 30.7        | 29.7        | 30.6        | 20.3        |
| Sigma A coordinate error (Å <sup>2</sup> )   | 0.09        | 0.17        | 0.21        | 0.29        | 0.14        |

*R*<sub>sym</sub> = (Σ(|*I* - <*I*>|)/Σ(*I*)), where <*I*> is the average intensity of multiple measurements. *R*<sub>cryst</sub> and *R*<sub>free</sub> = (Σ|*F*<sub>o</sub> - *F*<sub>c</sub>|)/(Σ|*F*<sub>o</sub>|). *R*<sub>free</sub> was calculated over 5% of the amplitudes not used in refinement. r.m.s., root mean square; pre-IS, pre-insertion site; post-IS, post-insertion site. See text for definition.

\*Values in parentheses correspond to those in the outer resolution shell.

†Data collected using Rigaku rotating anode X-ray generator.

‡Data collected at beamline 22-ID, APS.



**Figure 4** Mutagenic translesion replication of 8oxoG by BF in crystals. **a**, **b**, Structures of mutagenic 8oxoG replication (pink) are superimposed with structures of normal replication (grey). Before dATP incorporation, 8oxoG occupies the pre-insertion site as shown in Fig. 3b. After dATP incorporation, the A-8oxoG mismatch mimics a cognate A-T base pair at the post-insertion site (**a**). No polymerase distortions are observed after extension of the A-8oxoG mismatch (**b**). **c**, Hoogsteen base pairing between dA and 8oxoG with simulated annealing omit electron density (contoured at 4 $\sigma$ ). A-8oxoG base pair parameters, with normal parameters for the A-T base pair in parentheses, are shown below.

DNA mismatch that is treated by the polymerase as a normal base pair.

After extension and translocation to the DNA duplex binding region, the A-8oxoG mismatch continues to mimic a cognate base pair, failing to distort the polymerase active site and evading error detection (Fig. 4b). Distortions are also absent after extension of C-8oxoG (Fig. 3d), even though this base pair resembles a DNA mismatch at the post-insertion site. After extension of C-8oxoG, template distortion induced by internal steric strain of the *anti*-

8oxoG persists (Fig. 3d, inset), but the distortion is now localized to the site of the base pair and is not transmitted back to the polymerase active site, as are mismatches in undamaged DNA. The absence of long-range distortions is attributable to the stable Watson-Crick or Hoogsteen base-pairing configurations in C-8oxoG and A-8oxoG respectively, whereas base-pairing configurations of mismatches in undamaged heteroduplex DNA are less stable. Distortions induced by mismatch extension represent the final opportunity for the polymerase to recognize a DNA mismatch. By mimicking a cognate base pair, the A-8oxoG mismatch also evades this final error recognition mechanism.

Preferential formation of the A-8oxoG Hoogsteen base pair over the stable C-8oxoG base pair is enabled by guanine oxidation, a process that alters the hydrogen-bonding configurations of the oxidized base. The presence of 8oxoG inverts the molecular logic of mismatch recognition: the correct C-8oxoG base pair is recognized as a mismatch, whereas the mutagenic A-8oxoG base pair mimics the structure of a cognate base pair. Consequently, the incorporation efficiency of C-8oxoG is much less than that of A-8oxoG, accounting for the mutagenic properties of this lesion. This mechanism of error-prone replication of 8oxoG suggests that the mutagenic potential of other types of DNA damage depends in part on their ability to adopt alternative modes of base pairing with mismatched bases that evade polymerase error detection mechanisms. □

## Methods

### Preparation of protein and DNA

BF was purified as described<sup>23</sup>. DNA oligonucleotides containing 8oxoG (5'-CAT-8oxoG-CGAGTCAGGCT (template 1) and 5'-CTA-8oxoG-CGAGTCAGGCT (template 2)) were synthesized using UltraMILD CE phosphoramidites (trityl off mode). Template oligonucleotides and commercially synthesized complementary primer oligonucleotides were de-protected and purified by reversed-phase high-performance liquid chromatography (HPLC). Template oligonucleotides were analysed by HPLC and matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF).

### In vitro replication of 8oxoG by BF

The template 15-nucleotide DNA oligonucleotides used for crystallization and an unmodified 15-nucleotide DNA oligonucleotide identical in sequence but with guanine in place of 8oxoG were annealed to a 5' [<sup>32</sup>P]-labelled primer (5'-GCTGACTCG) at a molar ratio of 1:2. For semi-quantitative analyses (Fig. 2), primers were extended (standard 10  $\mu$ l reactions containing 50 mM Tris pH 8.0, 5 mM MgCl<sub>2</sub>, 1 mM DTT, 250  $\mu$ g ml<sup>-1</sup> BSA, 2.5% glycerol, 10 nM annealed duplex, 1 nM BF, room temperature) with 0.1–100  $\mu$ M dNTP (dCTP for accurate incorporation opposite 8oxoG; sequential addition of dCTP and dATP for accurate extension past 8oxoG; dATP for misincorporation opposite and accurate extension past 8oxoG). To determine incorporation preference opposite 8oxoG in a competitive reaction, dATP and dCTP were included at equimolar concentrations (100  $\mu$ M). Control reactions with dATP or dCTP mark the different mobilities of primers extended with dATP or dCTP. Quantitative primer extension reactions to determine nucleotide incorporation kinetics opposite 8oxoG and unmodified guanine used 100 nM duplex DNA with BF concentrations and reaction times optimized to ensure linearity and single hit kinetics (<20% of primers extended under steady-state conditions)<sup>17,24</sup>. Concentrations of the dATP or dCTP varied from 0 to 2 mM. Reactions were performed at 25 °C and terminated by addition of 10  $\mu$ l 95% formamide/10 mM EDTA and heating to 95 °C for 10 min. Primers and products were separated by electrophoresis (20% polyacrylamide/7 M urea) and quantified (Molecular Dynamics Storm 840 Phosphorimager). Michaelis constant ( $K_m$ ) and  $V_{max}$  values were determined by a least-squares nonlinear fit of the data to the Michaelis-Menten equation (Supplementary Data 1).

### Catalysis in the crystal and structure determination

Crystals of BF bound to DNA duplex containing 8oxoG at the pre-insertion site were grown as described<sup>9,11</sup>. Accurate incorporation of 8oxoG was observed in crystals of BF bound to template 1 soaked in a solution (60% saturated ammonium sulphate, 0.1 M Tris pH 8.0, and 2.5% MPD) with 30 mM dCTP. C-8oxoG was extended by soaking BF-DNA duplex crystals containing an enzymatically incorporated C-8oxoG base pair at the post-insertion site in a 30 mM dATP soak solution. Mutagenic replication of 8oxoG was observed in crystals of BF bound to template 2 with 8oxoG at the pre-insertion site soaked in 30 mM dATP (60% saturated ammonium sulphate, 0.1 M MES pH 5.8, and 2.5% MPD) or a 30 mM dATP/dTTP mixture (60% saturated ammonium sulphate, 0.1 M Tris pH 8.0, and 2.5% MPD) to observe incorporation or both incorporation and extension, respectively. Crystals were flash frozen in liquid nitrogen for storage and data collection (98 K on a RAXIS IV or RAXIS II detector; or APS beamline 22ID). Data were processed using HKL2000 (ref. 25). All crystals belong to space group  $P2_12_12_1$  and contain one molecule per asymmetric unit. Phases were calculated using molecular replacement

(Protein Data Bank entry 1L3S as the probe) and models refined using CNS<sup>26</sup>. CNS topology and parameter files for 8oxoG refinement were generated with PRODRG<sup>27</sup>. Structures were superimposed using the C $\alpha$  atoms of the palm subdomain (residues 646–655, 823–838, 863–869).

Received 28 May; accepted 2 August 2004; doi:10.1038/nature02908.  
Published online 22 August 2004.

- Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362**, 709–715 (1993).
- Finkel, T. & Holbrook, N. J. Oxidants, oxidative stress and the biology of ageing. *Nature* **408**, 239–247 (2000).
- Shibutani, S., Takeshita, M. & Grollman, A. P. Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG. *Nature* **349**, 431–434 (1991).
- Greenblatt, M. S., Bennett, W. P., Hollstein, M. & Harris, C. C. Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenesis. *Cancer Res.* **54**, 4855–4878 (1994).
- Hainaut, P. *et al.* IARC Database of p53 gene mutations in human tumors and cell lines: updated compilation, revised formats and new visualisation tools. *Nucleic Acids Res.* **26**, 205–213 (1998).
- Lowe, L. G. & Guengerich, F. P. Steady-state and pre-steady-state kinetic analysis of dNTP insertion opposite 8-oxo-7,8-dihydroguanine by *Escherichia coli* polymerases I exo- and II exo. *Biochemistry* **35**, 9840–9849 (1996).
- Furge, L. & Guengerich, F. P. Analysis of nucleotide insertion and extension at 8-oxo-7,8-dihydroguanine by replicative T7 polymerase exo- and human immunodeficiency virus-1 reverse transcriptase using steady-state and pre-steady-state kinetics. *Biochemistry* **36**, 6475–6487 (1997).
- Freisinger, E., Grollman, A. P., Miller, H. & Kisker, C. Lesion (in)tolerance reveals insights into DNA replication fidelity. *EMBO J.* **23**, 1494–1505 (2004).
- Johnson, S. J., Taylor, J. S. & Beese, L. S. Processive DNA synthesis observed in a polymerase crystal suggests a mechanism for the prevention of frameshift mutations. *Proc. Natl Acad. Sci. USA* **100**, 3895–3900 (2003).
- Johnson, S. J. & Beese, L. S. Structures of mismatch replication errors observed in a DNA polymerase. *Cell* **116**, 803–816 (2004).
- Kiefer, J. R., Mao, C., Braman, J. C. & Beese, L. S. Visualizing DNA replication in a catalytically active *Bacillus* DNA polymerase crystal. *Nature* **391**, 304–307 (1998).
- Goodman, M. F. Hydrogen bonding revisited: geometric selection as a principal determinant of DNA replication fidelity. *Proc. Natl Acad. Sci. USA* **94**, 10493–10495 (1997).
- Kool, E. T. Active site tightness and substrate fit in DNA replication. *Annu. Rev. Biochem.* **71**, 191–219 (2002).
- Miller, H. & Grollman, A. P. Kinetics of DNA polymerase I (Klenow fragment exo-) activity on damaged DNA templates: effect of proximal and distal template damage on DNA synthesis. *Biochemistry* **36**, 15336–15342 (1997).
- Krahn, J. M., Beard, W. A., Miller, H., Grollman, A. P. & Wilson, S. H. Structure of DNA polymerase beta with the mutagenic DNA lesion 8-oxodeoxyguanine reveals structural insights into its coding potential. *Structure (Camb.)* **11**, 121–127 (2003).
- Echols, H. & Goodman, M. F. Fidelity mechanisms in DNA replication. *Annu. Rev. Biochem.* **60**, 477–511 (1991).
- Goodman, M. F., Creighton, S., Bloom, L. B. & Petruska, J. Biochemical basis of DNA replication fidelity. *Crit. Rev. Biochem. Mol. Biol.* **28**, 83–126 (1993).
- Kunkel, T. A. & Bebenek, K. DNA replication fidelity. *Annu. Rev. Biochem.* **69**, 497–529 (2000).
- Nair, D. T., Johnson, R. E., Prakash, S., Prakash, L. & Aggarwal, A. K. Replication by human DNA polymerase- $\alpha$  occurs by Hoogsteen base-pairing. *Nature* **430**, 377–380 (2004).
- Ling, H., Boudsocq, F., Plosky, B. S., Woodgate, R. & Yang, W. Replication of a cis-syn thymine dimer at atomic resolution. *Nature* **424**, 1083–1087 (2003).
- Culp, S. J., Cho, B. P., Kadlubar, F. F. & Evans, F. E. Structural and conformational analyses of 8-hydroxy-2'-deoxyguanosine. *Chem. Res. Toxicol.* **2**, 416–422 (1989).
- Kouchakjian, M. *et al.* NMR structural studies of the ionizing radiation adduct 7-hydro-8-oxodeoxyguanosine (8-oxo-7H-dG) opposite deoxyadenosine in a DNA duplex. 8-Oxo-7H-dG(syn).dA(anti) alignment at lesion site. *Biochemistry* **30**, 1403–1412 (1991).
- Kiefer, J. R. *et al.* Crystal structure of a thermostable *Bacillus* DNA polymerase I large fragment at 2.1 Å resolution. *Structure* **5**, 95–108 (1997).
- Boosalis, M. S., Petruska, J. & Goodman, M. F. DNA polymerase insertion fidelity. Gel assay for site-specific kinetics. *J. Biol. Chem.* **262**, 14689–14696 (1987).
- Otwinowski, Z. & Minor, W. *Methods in Enzymology, Macromolecular Crystallography (Part A): Processing of X-ray Diffraction Data Collected in Oscillation Mode* (Academic Press, New York, 1997).
- Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- van Aalten, D. M. *et al.* PRODRG, a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules. *J. Comput. Aided Mol. Des.* **10**, 255–262 (1996).

**Supplementary Information** accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank H. W. Hellinga for critical reading of the manuscript. Research was carried out in part at the SER-CAT 22-ID beamline at the Advanced Photon Source (Argonne National Laboratory), which is supported by the US Department of Energy, Office of Energy Research. We also thank J. J. Warren for assistance with data collection. The work was supported by grants to L.S.B. from the National Cancer Institute and the Duke Comprehensive Cancer Center, and to T.C. from the Deutsche Forschungsgemeinschaft, the Volkswagen Stiftung and the Fonds of the German Chemical Industry. M.O. is supported by a fellowship from the Boehringer Ingelheim Foundation.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to L.S.B. (lsb@biochem.duke.edu). Coordinates have been deposited in the Protein Data Bank under accession codes 1U45, 1U47, 1U48, 1U49 and 1U4B.

## Structure of the ESCRT-II endosomal trafficking complex

Aitor Hierro<sup>1</sup>, Ji Sun<sup>2</sup>, Alexander S. Rusnak<sup>2</sup>, Jaewon Kim<sup>1</sup>, Gali Prag<sup>1</sup>, Scott D. Emr<sup>2</sup> & James H. Hurley<sup>1</sup>

<sup>1</sup>Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, US Department of Health and Human Services, Bethesda, Maryland 20892-0580, USA

<sup>2</sup>Department of Cellular and Molecular Medicine and Department of Chemistry and Biochemistry and Howard Hughes Medical Institute, University of California at San Diego, 9500 Gilman Drive, La Jolla, California 92093-0668, USA

The multivesicular-body (MVB) pathway delivers transmembrane proteins and lipids to the lumen of the endosome. The multivesicular-body sorting pathway has crucial roles in growth-factor-receptor downregulation<sup>1</sup>, developmental signalling<sup>2–4</sup>, regulation of the immune response<sup>5</sup> and the budding of certain enveloped viruses such as human immunodeficiency virus<sup>6</sup>. Ubiquitination is a signal for sorting into the MVB pathway<sup>7,8</sup>, which also requires the functions of three protein complexes, termed ESCRT-I, -II and -III (endosomal sorting complex required for transport)<sup>7,9,10</sup>. Here we report the crystal structure of the core of the yeast ESCRT-II complex, which contains one molecule of the Vps protein Vps22, the carboxy-terminal domain of Vps36 and two molecules of Vps25, and has the shape of a capital letter 'Y'. The amino-terminal coiled coil of Vps22 and the flexible linker leading to the ubiquitin-binding NZF domain of Vps36 both protrude from the tip of one branch of the 'Y'. Vps22 and Vps36 form nearly equivalent interactions with the two Vps25 molecules at the centre of the 'Y'. The structure suggests how ubiquitinated cargo could be passed between ESCRT components of the MVB pathway through the sequential transfer of ubiquitinated cargo from one complex to the next.

The MVB pathway delivers transmembrane proteins and lipids into small vesicles that invaginate into the lumen of the endosome. MVBs then fuse with the vacuole/lysosome and release the vesicles into the lumen where they are degraded by the hydrolases contained in the vacuole/lysosome<sup>1</sup>. Proteins that remain in the limiting membrane of the MVB are either delivered to the lysosomal/vacuolar limiting membrane or they are recycled to the Golgi complex or the plasma membrane. Therefore, the MVB sorting pathway plays a critical role in the trafficking of numerous cargo proteins within the endosomal membrane system.

Sorting of proteins into the MVB pathway is a complex, multi-step process. It requires coordinated functions of many protein complexes that are conserved from yeast to humans. Screens in *Saccharomyces cerevisiae* have identified numerous genes required for the sorting and trafficking of proteins into the MVB. Mutations in these genes result in defects in the sorting of transmembrane proteins, such as the G-protein-coupled receptor, Ste2p, from the plasma membrane to the vacuolar lumen by means of the MVB pathway. Thus far, more than 20 genes referred to as class E VPS (vesicular protein sorting) genes have been identified<sup>11,12</sup>.

During MVB sorting, Hrs–Vps27 is first recruited to the early endosome by virtue of its FYVE domain interaction with PI(3)P (refs 13–15) and its UIM (ubiquitin interacting motif) interaction with ubiquitinated cargo. It then recruits the ESCRT-I complex (composed of Vps23, 28, 37) to the membrane<sup>16</sup>. ESCRT-I recruits the downstream ESCRT-II and ESCRT-III complexes<sup>9</sup>. After the ESCRTs have been recruited to the endosome membrane, the AAA-type ATPase Vps4 binds ESCRT-III and following MVB vesicle formation catalyses the dissociation of the ESCRT protein complexes in an ATP-dependent manner for further rounds of protein sorting<sup>17</sup>. The MVB sorting process is topologically equivalent to the



budding process for envelope viruses such as human immunodeficiency virus (HIV), equine infectious anaemia virus (EIAV) and Ebola, suggesting a potential role for the ESCRTs in this event. In fact, it has been shown that TSG101-Vps23p binds to the p6 domain of HIV-1 Gag protein and inhibition of this binding inhibits viral budding<sup>6</sup>. The same inhibitory effect is seen with dominant negative SKD1, the mammalian Vps4 orthologue<sup>6</sup>.

Here we report the 3.6 Å resolution structure of the quaternary ESCRT-II complex containing Vps22, Vps36 and two Vps25 molecules. Apart from the two NZF domains of Vps36, nothing is known about the structures of any of the ESCRT subunits alone or in the complex. These structures, in conjunction with mutational analyses, provide a framework for an initial understanding of the mechanisms for ESCRT assembly and function.

The 3.6 Å crystal structure of the ESCRT-II core complex (Supplementary Table 1 and Fig. 1a) contains two molecules of Vps25, one molecule of Vps22 and one copy of the C-terminal 171 amino acid residues of Vps36. The observed stoichiometry of 1 Vps22: 2 Vps25: 1 Vps36 leads to a calculated molecular weight of 138 kDa for the intact complex. This is consistent with the 155 kDa estimated on the basis of size exclusion chromatography of ESCRT-II purified from yeast<sup>9</sup>. The complex adopts the shape of a capital letter 'Y', with overall dimensions of 120 × 85 × 52 Å (Fig. 1b). The base of the 'Y' consists of one of the two molecules of Vps25; one of the branches of the 'Y' consists of the second molecule of Vps25, and the other branch is formed by a subcomplex consisting of Vps22 and the C-terminal domain of Vps36. The N terminus of Vps22 is a single α-helix that protrudes away from the tip of the 'Y' shape.

Vps22 consists of a single N-terminal coiled coil followed by two winged-helix (WH) domains (Supplementary Fig. 1a). All but the first 19 residues are ordered. The N-terminal domain consists of helices H2–H4, the C-terminal half of H1 and the N-terminal half of H5. Helix H5 is shared between the helical domain and the first WH domain. WH domains take the form HSHHSWSW, where H denotes a helix, S a β-strand and W a "wing"<sup>18</sup>, a loop inserted between or after the latter two of the three β-strands. Thus WH1 of Vps22 consists of H5–H7 and S1–S3. Strands 1–3 form an anti-parallel β-sheet with strand order S1–S3–S2. A one-turn helix follows WH2 and immediately precedes the C terminus.

The first helix projects 45 Å away from the core and forms a dimeric coiled coil with a symmetry-related Vps22 molecule. The presence of a coiled coil in this region was anticipated on the basis of sequence analysis<sup>9</sup>. We do not observe dimerization of ESCRT-II complexes isolated from yeast<sup>9</sup> or in solution (not shown), and ESCRT-II dimerization is therefore unlikely to be the normal function of the Vps22 coiled coil. All of the ESCRT-III subunits

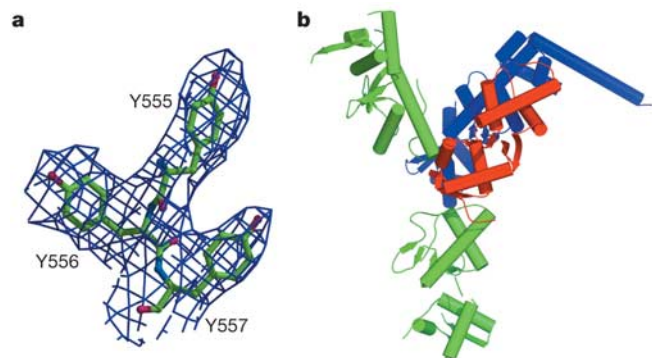
contain predicted coiled coils. On the basis of predictions from sequence analysis, it was proposed that an interaction occurs between a Vps22 coiled coil and coiled coils in the ESCRT-III subunits<sup>9</sup>.

Vps25 is the most flexible of the three subunits. Flexible internal loops between residues 56–70, 110–115, 146–149 and 162–164, as well as the C-terminal ten residues, could not be visualized in electron density. The N terminus of Vps25 consists of linked α- and 3<sub>10</sub>-helical turns (Supplementary Fig. 1b). The diproline units Pro-5 and -6, and Pro-11 and -12, respectively, initiate the two α-helical turns. A modified WH motif follows immediately after the linked turns. The Vps25 WH1 is missing the middle helix, and contains four additional β-strands inserted in its place.

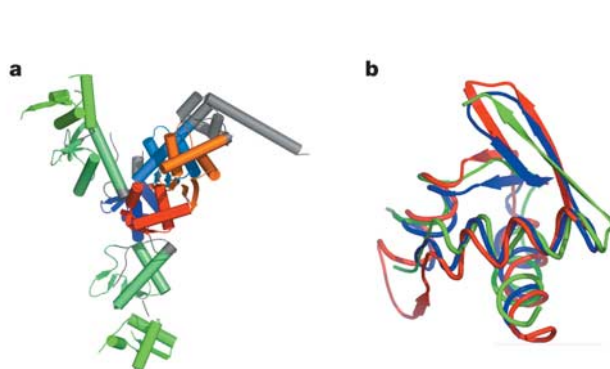
The Vps36 C-terminal domain consists of just two canonical WH domains (Supplementary Fig. 1c). There are short N- and C-terminal extensions and a short loop linking the two domains. Crystallization trials were initiated with a mixture of intact Vps22, Vps25 and Vps36. Upon microsequencing of redissolved crystals, it was determined that the crystallized complex contained a fragment of Vps36 comprising residues 396–566. The cleavage is attributed to trace amounts of protease contamination in the crystallized material. The proteolytic stability of the C-terminal portion of Vps36 is consistent with the structural finding that it makes extensive interactions within the core of the complex. The N-terminal portion of Vps36 contains two copies of a NZF zinc-finger motif. The second NZF finger of Vps36 binds ubiquitin and is thought to be the main site of interaction between ESCRT-II and ubiquitinated cargo proteins<sup>19</sup>. The proteolytic susceptibility of the region connecting the second NZF motif at residue 203 and the start of the C-terminal domain at 396 suggests that a flexible linker connects these two domains.

None of the three ESCRT-II subunits has discernable homology between each other at the amino acid sequence level. However all three have in common a structure based on two copies of the WH fold (Fig. 2a). For example, the WH2 of Vps36 superimposes on the WH2 of Vps22 with a root mean square deviation (r.m.s.d.) of 1.8 Å over 59 Cα positions (Fig. 2b). The internal WH folds also superimpose well on each other. The WH1 and WH2 folds of Vps22 can be superimposed with a r.m.s.d. of 1.9 Å for 51 Cα positions. In contrast to the similarity in the individual WH folds, the relative orientation of the two WH folds in Vps22 and Vps36 differ markedly, and the tandem WH fold cannot be superimposed simultaneously.

The Vps22–Vps36 interaction is the most extensive in the complex (Fig. 3a, b). This interaction buries a total of 2,527 Å<sup>2</sup> of solvent-accessible surface area from both partners, over a contact area whose longest dimension extends to 53 Å. The two subunits form a roughly parallel side-to-side arrangement, with the WH1



**Figure 1** Structure of the ESCRT-II complex. **a**, Electron density from solvent-flattened SIRAS map contoured at 0.7 σ in the vicinity of the Vps25–Vps36 interface showing the refined structure in a stick model for Vps36 Tyr 555, Tyr 556 and Tyr 557. **b**, Overall structure of the complex: Vps22 is blue, Vps36 red and Vps25 green.



**Figure 2** Winged helix folds in ESCRT-II. **a**, WH folds are highlighted in light and dark blue (Vps22 WH1 and WH2), light and dark green (Vps25 WH1 and WH2), and orange and red (Vps36 WH1 and WH2); the rest of the structure is coloured grey. **b**, Superposition of WH2 of Vps22, WH2 of Vps36 and WH2 of Vps25. Colours are as in Fig. 1.

and WH2 folds each in contact with their cognates from the opposing subunit.

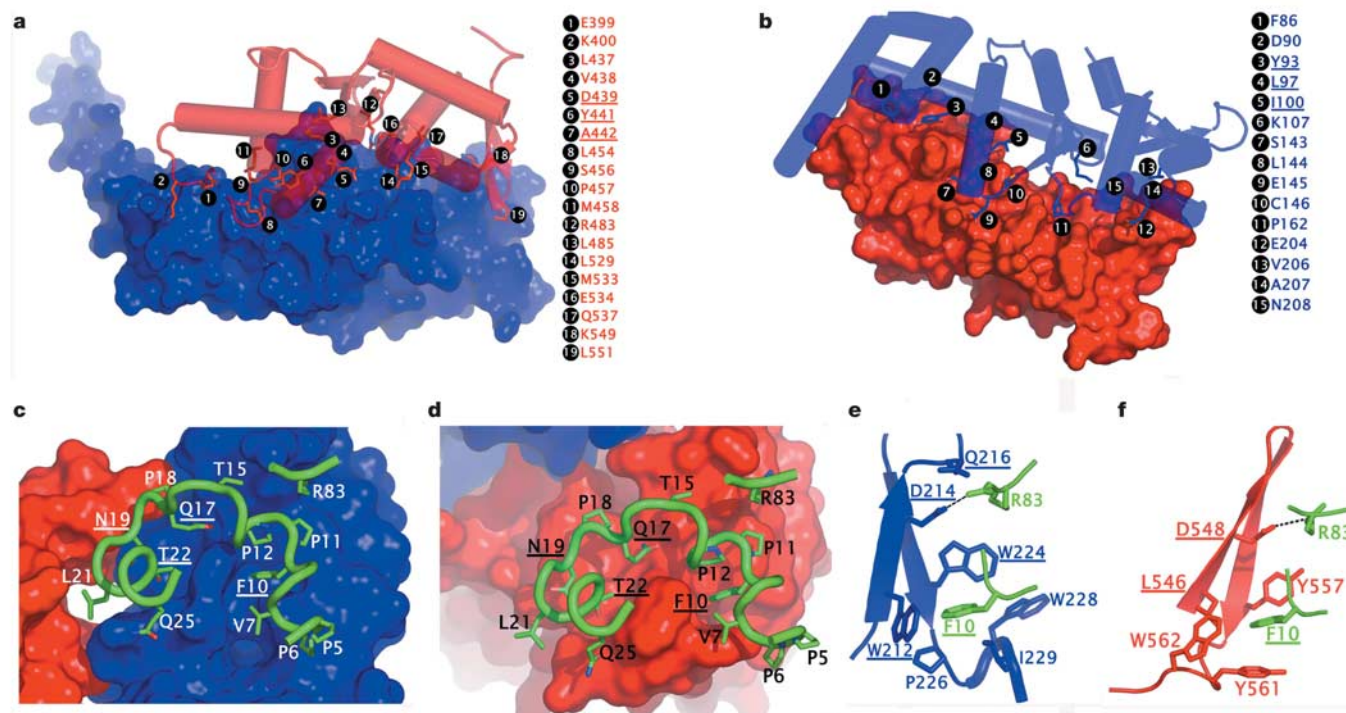
Vps22 uses its C-terminal sheet and wings and adjacent regions to bind one Vps25 molecule (Fig. 3c). Vps25 interacts with Vps22 through its N terminus, helix H1 and the loop immediately preceding H3. The interaction buries a total of 1,689 Å<sup>2</sup> of solvent-accessible surface area from both partners. The focal point of the Vps25 binding site on Vps22 is an aromatic cage made up of the side chains of Trp 212, Trp 224 and Trp 228. Pro 226 and Ile 229 also contribute to the walls of the cage. Together these residues completely surround and bury Phe 10 of Vps25. The hydrophobic contact surface extends over a radius of roughly 10 Å surrounding Phe 10. In addition to the many hydrophobic interactions, a salt bridge is formed by the well-ordered side chains of Asp 214 (Vps22) and Arg 83 (Vps25). Vps22 approaches within 4 Å of Vps25 molecule 2, which is tightly bound to Vps36. Vps25 molecule 2 has little interaction with Vps22, burying a total of only 332 Å<sup>2</sup>.

The parallels between the Vps36–Vps25 and Vps22–Vps25 interactions are striking (Fig. 3c–f). Phe 10 of Vps25 is buried in an aromatic cage on Vps36. The cage on the sheet and wings of Vps36 is formed by Leu 546, Tyr 557, Tyr 561 and Trp 562. A salt bridge is formed between Arg 83 of Vps25 and Asp 548 of Vps36. The total surface area buried is 1,524 Å<sup>2</sup>, identical to that between Vps22 and Vps25, within error. In contrast to Vps22, Vps36 has significant interactions with the Vps25 molecule 1. This secondary interaction buries 736 Å<sup>2</sup> of surface area. Helix H5 and the hairpin turn between S5 and S6 of Vps36 make interactions with the N-terminal part of H1 of Vps25 molecule 1.

To determine which subunit interfaces participate in the assembly of ESCRT-II, and which interactions are required for function, mutations were engineered that blocked each of the subunit interfaces (Fig. 4). The Vps36 mutation Y441A disrupts hydrophobic

interactions between Vps36 and Leu 97, Ile 100, Leu 141 and Leu 144 of Vps22 (Fig. 3a). Vps36 Y441A has a class E phenotype, and forms a 1:1 Vps25:Vps36 subcomplex (Supplementary Table 2). The Vps22 mutation I100D disrupts the same hydrophobic cluster at Y441A (Fig. 3b). This mutant has a class E phenotype and forms a 1:1 Vps22:Vps25 subcomplex (Fig. 4a). The behaviour of these mutants shows that the Vps22–Vps36 interface is subject to mutational disruption, that subcomplexes can exist *in vivo* without this interface and that subcomplexes lacking Vps22 or Vps36 are not functional *in vivo*.

To analyse the functions of the two Vps25 molecules, mutations were made in Vps25 and in the binding sites for Vps25 on Vps22 and Vps36. The mutation F10D at the heart of the Vps25 interface with both of the other two subunits (Fig. 3c–f) leads to a class E phenotype and the detection of Vps25 as a monomer only. The mutation R83D, which removes salt bridges with Vps22 Asp 214 and Vps36 Asp 548 (Fig. 3c–f), also leads to a class E phenotype and a monomer Vps25 (Fig. 4a, b). Of the three mutations in the Vps22 binding site for Vps25, W212D, D214A and Q216A, the first two lead to class E phenotypes and formation of 1:1:1 Vps22:Vps25:Vps36 subcomplexes (Fig. 4a, b). The mutation at Gln 216, which is more solvent-exposed and appears to make weaker interactions than the other two residues mutated in this site, has a wild-type sorting phenotype and forms the wild-type 1:2:1 tetramer. The Vps36 mutation D548A is the structural counterpart of the Vps22 mutation D214A, but has a wild-type phenotype. Vps25 molecule 2, the partner for the Vps36 binding site, buries 250 Å<sup>2</sup> more in interactions with the Vps22–Vps36 subcomplex than does Vps25 molecule 1. The additional interaction area may make it more difficult to disrupt this interaction. The mutation D548R disrupts the interface to a greater degree than D548A, and results in a class E phenotype and a 1:1:1



**Figure 3** ESCRT-II complex assembly and subunit interactions. **a**, Vps22 (solid; blue) interactions with Vps36 (ribbons, cylinders and sticks; red). **b**, Vps36 (solid; red) interactions with Vps22 (ribbons, cylinders and sticks; blue). **c**, First Vps25 subunit (C-chain) N terminus (stick model; green) bound to surface of Vps22–Vps36 subcomplex (Vps22, blue; Vps36, red). **d**, Second Vps25 (D-chain) N terminus (stick model; green)

bound to subcomplex (surface Vps36, red; Vps22, blue). **e**, Wings and sheet of Vps22 (blue) interacting with Vps25 molecule 1 (green). **f**, Wings and sheets of Vps36 (red) interacting with Vps25 molecule 2 (green). Underlined residues were mutated as indicated in Supplementary Table 2.



Vps22:Vps25:Vps36 subcomplex (Fig. 4a). These results demonstrate that both Vps25 molecules are required for ESCRT-II complex assembly and sorting function.

The structure of the salt bridge between Arg 83 of Vps25 and Asp 548 of Vps36 suggested that a charge reversal mutation at one site might serve as a second-site suppressor of a mutation in the other site. The double mutant Vps25<sup>R83D</sup> Vps36<sup>D548R</sup> has a class E phenotype and a combination of Vps25 monomers at 1:1:1 Vps22:Vps25:Vps36 subcomplex is observed (Fig. 4a). The class E phenotype is consistent with the abrogation of the Vps22–Vps25 interface by the Vps25 R83D mutation, which is not rescued by the Vps36 D548R mutation. The existence of a 1:1:1 Vps22:Vps25:Vps36 subcomplex, as compared to its absence in

the R83D single mutant, is consistent with the rescue of the Vps25–Vps36 interface by the second mutation. This result elegantly corroborates the structural observation of charge complementarity in the Vps25 binding sites.

The study of the ESCRT-II complex reported here offers structural insight into the core complexes that sort ubiquitinated membrane proteins into multivesicular bodies. The central observations are that ESCRT-II consists of a rigid, Y-shaped core scaffold with a protruding coiled coil for putative interactions with other ESCRT complexes, and a ubiquitin-binding domain attached by a long flexible linker. The mechanism of ubiquitinated cargo sorting can now be considered in the context of three-dimensional structural information. This mechanism is operative in human cells and is of profound medical importance. The budding or fission of HIV-1 and other retroviruses from human macrophages, and under appropriate conditions other cell types, occurs by this pathway. Our report concerns the yeast ESCRT-II complex; the subunits of yeast ESCRT-II are highly homologous (Supplementary Fig. 2) with their human counterparts throughout the core (although the human homologue of Vps36 lacks the N-terminal NZF domains), and therefore the structure reported here provides a reasonable basis for insights into the sorting mechanism in both yeast and human cells.

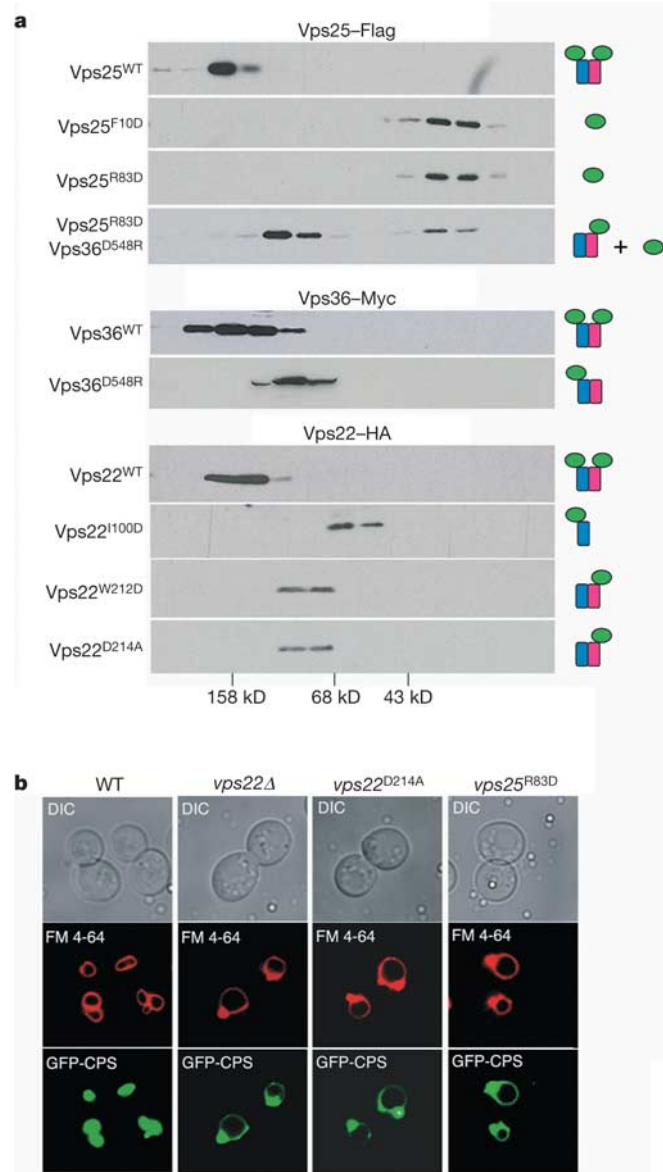
At least three ubiquitin-binding domains are involved in three sequential steps in the recruitment and hand-off of ubiquitinated cargo between the Vps27–Hse1 (Hrs–STAM in mammals) and ESCRT complexes. The Vps27–Hse1 complex recruits ubiquitinated cargo through the UIMs in Vps27. The cargo is transferred to ESCRT-I where it is bound to by the UEV (ubiquitin conjugating enzyme E2 variant) domain of Vps23. The cargo is then transferred to the second NZF domain of Vps36 in ESCRT-II. The affinity of each of these domains for ubiquitin is modest: ~250, 500 and 180  $\mu$ M for the Vps27 UIM-1<sup>20,21</sup>, Vps23 UEV<sup>6,22,23</sup> and Vps36 NZF-2<sup>19</sup>, respectively. Because the affinities are low and do not follow an obvious hierarchy, it is difficult to account for the specificity and directionality of trafficking on the basis of affinities alone.

The ESCRT-II structure shows that the coiled coil domain is closely juxtaposed with the Y-shaped core of the complex. This suggests that the larger oligomers of ESCRT complexes could form a highly organized scaffold. Such a scaffold could prevent the ubiquitinated cargo from diffusing away from one of the low-affinity binding domains in the complex. In contrast to the close juxtaposition of the coiled coil with the core, the ubiquitin-binding domain is attached to the core by a long tether. Whereas the structure of the ESCRT-I complex is unknown, the situation may be similar in that there is a predicted 40-amino-acid flexible linker between the ubiquitin-binding UEV domain of Vps23 and the remainder of the subunit. This suggests that these complexes are equipped with long swinging arms for the transfer of cargo over distances of tens to hundreds of Å in the course of sorting (Supplementary Fig. 3). A complete mechanistic understanding of this pathway will require additional structures of the remaining complexes as well as those of the higher order complexes of complexes. The resolution of the first one of these structures provides the first detailed glimpse into the mechanism of ESCRT complex formation and a conceptual and practical foundation for the long-term effort in the field to reach this goal. □

## Methods

### Protein expression and purification

The DNA sequences encoding Vps22, Vps25 and Vps36 were amplified by polymerase chain reaction (PCR) to generate cassettes containing the Shine–Delgarno translational start signal and cloned directly into the polycistronic pST39 expression vector<sup>24</sup>. Vps22 was tagged with an N-terminal hexahistidine tag and a TEV protease cleavage site; the other subunits were untagged. The plasmid was transformed into *Escherichia coli* strain BL21(DE3) Codon Plus RIL and expressed by inducing with 0.5 mM isopropylthiogalactoside (IPTG) at an absorbance at 600 nm ( $A_{600}$ ) of 0.8 and grown at 15 °C for 36 h. Cells were lysed by high pressure homogenization in 50 mM Tris (pH 7.5), 300 mM NaCl, 5 mM  $\beta$ -mercaptoethanol (BME) and 100  $\mu$ l of protease inhibitor cocktail



**Figure 4** Mutational analysis of ESCRT-II subunit interfaces. **a**, Mutations at the molecular interface of ESCRT-II disrupt complex assembly, leading to subcomplex and monomer formation, as determined by gel filtration chromatography. Cartoons to the right (Vps25, green; Vps22, blue; and Vps36, red) indicate the corresponding complex/subcomplex composition of ESCRT-II and elution of molecular weight standards are indicated at the bottom. **b**, ESCRT-II subunit interface mutants exhibit defects in sorting MVB cargo (GFP–CPS) to the lumen of the vacuole and instead traffic them to the limiting membrane (FM4-64) of the vacuole.



(Sigma) per litre of culture medium. The ESCRT-II complex was isolated using Talon Co<sup>2+</sup> affinity chromatography (Clontech). The histidine tag was removed with TEV protease, concomitant with dialysis in 50 mM Tris (pH 8.0), 150 mM NaCl and 5 mM  $\beta$ ME. The ESCRT-II complex was further purified with a second Talon affinity column and a Superdex S 200 gel filtration column.

### Crystallization and data collection

Crystals of the ESCRT-II complex were obtained using hanging-drop vapour diffusion by mixing 3  $\mu$ l of 10 mg ml<sup>-1</sup> protein solution and 3  $\mu$ l of 100 mM HEPES (pH 7.4) and 1.4 M NaAcetate. Crystals grew to 0.7 mm per side in ~3 weeks. The content of the crystals was determined by mass spectrometry (MALDI-TOF) and automated N-terminal sequencing. Crystals were cryo-protected in mother liquor supplemented with 25% glycerol and flash-frozen in liquid nitrogen. Xenon-derivatized crystals were prepared in an Xcell pressure chamber (Oxford Cryosystems) at 20 bar for 5 min. Diffraction data were collected at  $\lambda = 1.0 \text{ \AA}$  from native and Xe-derivatized ESCRT-II crystals at beamline ID-22 at the Advanced Photon Source, using 0.5° oscillations, 1 s exposures and a Mar225 CCD (charge-coupled device) detector. Additional native data were collected at  $\lambda = 1.8 \text{ \AA}$  at beamline 9-2 at the Stanford Synchrotron Radiation Laboratory. Data were indexed and reduced using HKL2000 (HKL Research).

### Structure determination

Five Xe sites were located in an automated Patterson search using the program SOLVE<sup>25</sup>. The resulting density synthesis was improved by density modification in RESOLVE<sup>26</sup> using a solvent fraction of 0.75. The map was interpreted manually using the interactive graphics program O<sup>27</sup>. Given the low resolution of the density map, the assignment of an amino acid sequence to the subunits was greatly facilitated by the use of secondary structure predictions<sup>28</sup>. Subsequent to sequence assignment and model building, a  $\lambda = 1.8 \text{ \AA}$  native anomalous difference Fourier ( $\Delta F$ ) synthesis at 4.5  $\text{\AA}$  resolution was used to verify the locations of sulphur atoms in Cys and Met residues. The  $\Delta F$  synthesis contained peaks for 8 of 11 S atoms in Vps22 (Met 36, Met 66 and Cys 67 absent), 9 of 9 S atoms in Vps36, and 3 of 3 S atoms in the N-terminal half of Vps25. The  $\Delta F$  synthesis could not be interpreted reliably for the C-terminal half of Vps25 and the sequence assignment is therefore tentative for this region and for the N-terminal ~50 residues of Vps22. The model was refined in the CNS program using torsional dynamics and the maximum likelihood target function<sup>29</sup>. The final model has no residues in disallowed regions of the Ramachandran plot.

### Plasmid construction and yeast strains

DNA encoding a haemagglutinin (HA), Myc and Flag epitope were introduced just before the stop codon in the chromosomal copy of *VPS22*, *VPS36* and *VPS25*, respectively. *VPS22-HA*, *VPS36-Myc* and *VPS25-Flag* sequences were then amplified from genomic DNA. The *SalI/SpeI*-digested PCR product of *VPS22-HA*, the *SphI/SacII*-digested PCR product of *VPS36-Myc* and the *SalI/SacII*-digested PCR product of *VPS25-Flag* were ligated with *SalI/SpeI*-digested pRS416, *SphI/SacII*-digested pSJ001 and *SalI/SacII*-digested pRS416, respectively, to generate pMB170, pSJ004 and pSJ072. Vps22, Vps36 and Vps25 interaction site mutations were introduced into pMB170, pSJ004 and pSJ072 by Quickchange mutagenesis (Stratagene). pGO45 (green fluorescent protein (GFP)-carboxypeptidase S (CPS) (GFP-CPS)) has been described<sup>30</sup>. The following yeast strains were used: SEY6210 (*MAT $\alpha$  leu2-3, 112 ura3-52 his3- $\Delta$ 200 trp1- $\Delta$ 901 lys2-801 suc2- $\Delta$ 9*); MBY31 (SEY6210.1; *vps22 $\Delta$ 1::HIS3*); MBY30 (SEY6210; *vps36 $\Delta$ 1::HIS3*)<sup>31</sup>; BWY101 (SEY6210; *vps25 $\Delta$ 1::HIS3*); SJY021 (SEY6210; *vps25 $\Delta$ 1::HIS3 vps36 $\Delta$ 1::HIS3*).

### Microscopy

Living cells expressing the GFP-CPS chimera were harvested at an  $A_{600}$  of 0.4–0.6, labelled with FM4-64 for vacuolar membrane staining and resuspended in medium for visualization. Visualization of cells was performed on a fluorescence microscope (Axiovert S1002TV; Carl Zeiss MicroImaging, Inc.) equipped with fluorescein isothiocyanate (FITC) and rhodamine filters, captured with a digital camera (Photometrix) and deconvolved using Delta Vision software (Applied Precision Inc.). Results presented were based on observations of >120 cells.

### Gel filtration analysis

For gel filtration analysis, yeast cells were spheroplasted and lysed in PBS containing 0.1 mM AEBSE, 1  $\mu$ g ml<sup>-1</sup> pepstatin A, 1  $\mu$ g ml<sup>-1</sup> leupeptin, 1 mM benzamide and protease inhibitor cocktail (Complete; Roche Molecular Biochemicals). The lysate was precleared for 5 min at 300 g followed by a 100,000 g centrifugation. The following lysate was loaded onto a Sephacryl S200 column (16/60; Amersham Life Sciences) and separated with PBS. Fractions were analysed with western blotting using anti-HA, anti-Myc and anti-Flag antibodies.

Received 13 July; accepted 9 August 2004; doi:10.1038/nature02914.

Published online 25 August 2004.

1. Futter, C. E., Pearce, A., Hewlett, L. J. & Hopkins, C. R. Multivesicular endosomes containing internalized EGF-EGF receptor complexes mature and then fuse directly with lysosomes. *J. Cell Biol.* **132**, 1011–1023 (1996).
2. Deldrande, G. A., Lai, E. C. & Kintner, C. *Xenopus* neuralized is a ubiquitin ligase that interacts with XDelta1 and regulates Notch signaling. *Dev. Cell* **1**, 795–806 (2001).

3. Lai, E. C., Deldrande, G. A., Kintner, C. & Rubin, G. M. *Drosophila* neuralized is a ubiquitin ligase that promotes the internalization and degradation of Delta. *Dev. Cell* **1**, 783–794 (2001).
4. Pavlopoulos, E. *et al.* neuralized encodes a peripheral membrane protein involved in Delta signaling and endocytosis. *Dev. Cell* **1**, 807–816 (2001).
5. Kleijmeier, M. *et al.* Reorganization of multivesicular bodies regulates MHC class II antigen presentation by dendritic cells. *J. Cell Biol.* **155**, 53–63 (2001).
6. Garrus, J. E. *et al.* Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding. *Cell* **107**, 55–65 (2001).
7. Katzmman, D. J., Babst, M. & Emr, S. D. Ubiquitin-dependent sorting into the multivesicular body pathway requires the function of a conserved endosomal protein sorting complex, ESCRT-I. *Cell* **106**, 145–155 (2001).
8. Reggiori, F. & Pelham, H. R. B. Sorting of proteins into multivesicular bodies: ubiquitin-dependent and -independent targeting. *EMBO J.* **20**, 5176–5186 (2001).
9. Babst, M., Katzmman, D. J., Snyder, W. B., Wendland, B. & Emr, S. D. Endosome-associated complex, ESCRT-II, recruits transport machinery for protein sorting at the multivesicular body. *Dev. Cell* **3**, 283–289 (2002).
10. Babst, M., Katzmman, D. J., Estepa-Sabal, E. J., Meerloo, T. & Emr, S. D. ESCRT-III: An endosome-associated heterooligomeric protein complex required for MVB sorting. *Dev. Cell* **3**, 271–282 (2002).
11. Raymond, C. K., Roberts, C. J., Moore, K. E., Howald, I. & Stevens, T. H. Biogenesis of the vacuole in *Saccharomyces cerevisiae*. *Int. Rev. Cytol.* **139**, 59–120 (1992).
12. Katzmman, D. J., Odorizzi, G. & Emr, S. D. Receptor downregulation and multivesicular-body sorting. *Nature Rev. Mol. Cell Biol.* **3**, 893–905 (2002).
13. Katzmman, D. J., Stefan, C. J., Babst, M. & Emr, S. D. Vps27 recruits ESCRT machinery to endosomes during MVB sorting. *J. Cell Biol.* **162**, 413–423 (2003).
14. Bache, K. G., Brech, A., Mehlum, A. & Stenmark, H. Hrs regulates multivesicular body formation via ESCRT recruitment to endosomes. *J. Cell Biol.* **162**, 435–442 (2003).
15. Bilodeau, P. S., Urbanowski, J. L., Winistorfer, S. C. & Piper, R. C. The Vps27p-Hse1p complex binds ubiquitin and mediates endosomal protein sorting. *Nature Cell Biol.* **4**, 534–539 (2002).
16. Lu, Q., Hope, L. W. Q., Brach, M., Reinhard, C. & Cohen, S. N. TSG101 interaction with HRS mediates endosomal trafficking and receptor down-regulation. *Proc. Natl Acad. Sci. USA* **100**, 7626–7631 (2003).
17. Babst, M., Wendland, B., Estepa, E. J. & Emr, S. D. The Vps4p AAA ATPase regulates membrane association of a Vps protein complex required for normal endosome function. *EMBO J.* **17**, 2982–2993 (1998).
18. Gajiwala, K. S. & Burley, S. K. Winged helix proteins. *Curr. Opin. Struct. Biol.* **10**, 110–116 (2000).
19. Alam, S. L. *et al.* Ubiquitin interactions of NZF zinc fingers. *EMBO J.* **23**, 1411–1421 (2004).
20. Fisher, R. D. *et al.* Structure and ubiquitin binding of the ubiquitin-interacting motif. *J. Biol. Chem.* **278**, 28976–28984 (2003).
21. Swanson, K. A., Kang, R. S., Stamenova, S. D., Hicke, L. & Radhakrishnan, I. Solution structure of Vps27 UIM-ubiquitin complex important for endosomal sorting and receptor downregulation. *EMBO J.* **22**, 4597–4606 (2003).
22. Sundquist, W. I. *et al.* Ubiquitin recognition by the human TSG101 protein. *Mol. Cell* **13**, 783–789 (2004).
23. Teo, H., Vepritsiev, D. B. & Williams, R. L. Structural insights into ESCRT-I recognition of ubiquitinated proteins. *J. Biol. Chem.* in the press.
24. Tan, S. A modular polycistronic expression system for overexpressing protein complexes in *Escherichia coli*. *Protein Expr. Purif.* **21**, 224–234 (2001).
25. Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **55**, 849–861 (1999).
26. Terwilliger, T. C. Maximum-likelihood density modification. *Acta Crystallogr. D Biol. Crystallogr.* **56**, 965–972 (2000).
27. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron-density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
28. Rost, B. & Liu, J. F. The PredictProtein server. *Nucleic Acids Res.* **31**, 3300–3304 (2003).
29. Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54**, 905–921 (1998).
30. Odorizzi, G., Babst, M. & Emr, S. D. Fab1p PtdIns(3)P 5-kinase function essential for protein sorting in the multivesicular body. *Cell* **95**, 847–858 (1998).

**Supplementary Information** accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank P. P. di Fiore, M. Babst and B. Wendland for helpful discussions; M. Tabuchi for the Flag tag construct; and we acknowledge the use of the SER-CAT beamline at the APS, ANL and beamline 9-2 at the Stanford Synchrotron Radiation Laboratory. Use of the Advanced Photon Source was supported by the US Department of Energy, Basic Energy Sciences, Office of Science. The Stanford Synchrotron Radiation Laboratory, a national user facility operated by Stanford University on behalf of the US Department of Energy, is supported by the Department of Energy, Office of Biological and Environmental Research, and by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program and the National Institute of General Medical Sciences. S.D.E. is supported as an Investigator of the Howard Hughes Medical Institute.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to J.H.H. ([hurley@helix.nih.gov](mailto:hurley@helix.nih.gov)). Coordinates have been deposited with the Protein Data Bank with accession number 1U5T.

## Contacts

**Publisher:** Ben Crowe

**Editor:** Paul Smaglik

**Marketing Manager:** David Bowen

### European Head Office, London

The Macmillan Building  
4 Crinan Street  
London N1 9XW, UK  
Tel +44 (0) 20 7843 4961  
Fax +44 (0) 20 7843 4996  
e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

### Naturejobs Sales Director:

Nevin Bayoumi (4978)

### UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

### Scandinavia/ Spain/ Portugal:

Evelina Rubio Håkansson (4973)

**Natureevents:** Silke Opstrup (4994)

### France/ Switzerland:

Amelie Pequignot (4974)

### Production Manager:

Billie Franklin  
To send materials use London  
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

### Naturejobs web development:

Tom Hancock

### Naturejobs online production:

Stefan Hales

### European Satellite Office

**Germany/ Austria/ Italy/**

**The Netherlands/ Belgium:**

Patrick Phelan, Odo Wulffen

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mail: [p.phelan@nature.com](mailto:p.phelan@nature.com)

[o.wulffen@nature.com](mailto:o.wulffen@nature.com)

### US Head Office, New York

345 Park Avenue South,  
10th Floor, New York, NY 10010-1707  
Tel +1 800 989 7718  
Fax +1 800 989 7103  
e-mail: [naturejobs@natureny.com](mailto:naturejobs@natureny.com)

### US Sales Manager:

Peter Bless

### Japan Head Office, Tokyo

MG Ichigaya Building (5F),  
19-1 Haraikatamachi,  
Shinjuku-ku,  
Tokyo 162-0841  
Tel +81 3 3267 8751  
Fax +81 3 3267 8746

### Asia-Pacific Sales Director:

Rinoko Asami  
e-mail: [rasami@naturejpn.com](mailto:rasami@naturejpn.com)

# naturejobs

## The joy of the lab

Career talks usually feature speakers who discuss ways to get jobs outside the laboratory. There is good reason for this — long-term, well-funded research positions are hard to secure, and young scientists need to know the odds and the options. At a *Naturejobs* career symposium last month, held during the EuroScience Open Forum in Stockholm, Sweden, we featured a smorgasbord of scientists who have pursued work beyond the bench — including positions in the drug industry, communication and entrepreneurship.

But one speaker, who has stayed in the lab, reminded the young scientists in attendance that basic research is also worth pursuing and that, with determination, it can result in a satisfying career. Jon Storm-Mathisen of the University of Oslo's Centre for Molecular Biology and Neuroscience in Norway, offered some tips on how to succeed on this path.

One of the most important factors, says Storm-Mathisen, is a good mentor — someone who can inspire and generate an atmosphere of curiosity and drive. Another key point is to select a strong question to research. There are two potential pitfalls here, he notes. First, young scientists sometimes ask a question that is too easy. Initially, this might increase the odds of success, but the answer will often be empty and won't lead you forwards. Going out on a limb and asking a hard question can be frightening, but may result in greater long-term security. Second, young scientists are often tempted to follow the crowd — in terms of technology or disease area, for example. Doing so may help secure funding early on, but it won't help you to make your mark. "You shouldn't jump on any bandwagon," Storm-Mathisen says.

Finally, scientists need to have the right motivation. Storm-Mathisen uses one word: "Joy." In fact, that driver should be the same, regardless of where a scientist works.

**Paul Smaglik**  
*Naturejobs* editor



## Contents

### CAREER VIEW

#### Nuts & Bolts

Performance reviews

#### Graduate Journal

Try, try again

#### Movers

Elaine Ostrander

p228

### WWW.NATUREJOBS.COM

Career centre

Information on the  
scientific job market

#### FOCUS

#### SPOTLIGHT

#### RECRUITMENT

#### ANNOUNCEMENTS

#### EVENTS

## GRADUATE JOURNAL

## Try, try again

This summer I had the good fortune to attend a Gordon Research Conference on biomineralization — the way organisms create minerals and turn them into functional structures. The small size of the meeting and its 'off-the-record' policy promoted openness. And the interplay between junior and senior scientists — aided by group dining, multiple poster sessions and a few cold beverages at the end of the day — all but guaranteed frank discussions.

I was excited by the chance to meet successful colleagues from different backgrounds. From industry to academia, from conservative to fringe, no two successful researchers were obviously alike. Some junior attendees noted this wide diversity and wondered what could be the common denominator for success.

One reasonable candidate emerged: tenacity. None of these people seemed the kind that give up easily. My few years in research have not resulted in any easy successes. I can only imagine the number of failures that must be overcome during decades of dedicated effort. I have returned to the lab with new friends, inspiring stories, and a fresh relationship with my veritable forest of failures. I am excited to forge ahead until I too meet with some scientific success. ■

**Sidney Omelon is a PhD student in bone biomaterials at the Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Canada.**

## NUTS &amp; BOLTS

## Performance reviews

**L**ike attending a funeral or going to the dentist, annual performance reviews score highly as things to be avoided. Some scientists and researchers go to great lengths, often inadvertently, to get out of the process altogether by never setting a date for a review or by repeatedly rescheduling it until it vanishes completely. If this sounds vaguely familiar, it's time to adopt a new mindset and try a different approach. To get the most from your next review, take an active role, seek clarification and focus on development.

A classic mistake is to assume a passive 'let's see what happens' attitude, expecting your adviser or supervisor to do all of the preparing and presenting. You'll find that reviews work best when you play a proactive role and consider your value on an ongoing basis. Do the leg work throughout the year by



With Deb Koen  
Careers consultant

noting down your achievements. To avoid nasty surprises, seek feedback regularly. Try to anticipate your supervisor's areas of focus and be ready to discuss your contributions. If there are areas you've been trying to improve, cite your progress. Participate in the discussion by asking solid questions, actively listening, responding — directly but not defensively — and highlighting the key points you want to make.

Most importantly, you must seek clarification. Hearing that you're 'not doing well enough' isn't

constructive criticism because you can't act on vague generalities. You need to probe for details until you reach an understanding of the priorities and specific actions to take. Know what's expected and agree on a timetable.

Don't stop at your performance. Be sure to weave the topic of your development into the process. Although your development is ultimately your responsibility, your employer benefits as well. Set goals that align with those of your employer or sponsoring organization and with your own priorities for professional enrichment. Point out the value to your employer, and negotiate the resources you need.

By taking more control of the review process, you can both assess your progress and shape your future role. ■

**Deb Koen is vice-president of Career Development Services and a columnist for The Wall Street Journal's CareerJournal.com.**

## MOVERS

Elaine Ostrander, branch chief, cancer genetics, NHGRI, Bethesda, Maryland



**L**ike most scientists, Elaine Ostrander's work has benefited from collaborations. But somewhat unusually, she estimates that 90% of her publications feature work with scientists who were not based at her principal laboratory. In fact, she didn't even seek out those collaborators who have proved most valuable — they found her.

Ostrander feels that her career path was equally fortuitous. Her early thesis

work and postdoctoral research concentrated on DNA structure. But a cluster of meetings on the genomes of the fruitfly, nematode worm and mouse caught her attention and she decided to change tack. She moved to Lawrence Berkeley National Laboratory in 1991, just as it was beginning to map the dog genome, which offered her the opportunity she was seeking. Her good fortune continued when she met her key collaborators, who helped her forge links between dog and human.

She ran into her first major collaborator, Francis Galibert of the University of Rennes in France, in 1993 after she gave a talk at Cold Spring Harbor Laboratory. A little later, Janet Stanford, Kathleen Malone and Janet Daling from the Public Health Sciences division at the Fred Hutchinson Cancer Research Center in Seattle, Washington, approached her, saying that they had large human genetic sets for breast and

prostate cancer. These allowed the group to look for genetic mutations shared by both species in an effort to pinpoint genes related to disease.

Ostrander's guiding principle behind joining collaborations has remained constant. "Find people who are doing interesting things," she says, "and put your heads together to find ways to do things that are bigger than what you can each do on your own."

The need for bigger challenges is one reason why Ostrander is moving to the US National Human Genome Research Institute (NHGRI) in November. With the dog genome mapped, she is ready to take a wider look at gene function. Access to proteomics tools at the NHGRI's campus in Bethesda, Maryland, will allow her to ask broader questions. And the National Institute of Health's huge number of employees will provide her with plenty more potential collaborators. ■

**CV** **1993–2004:** Clinical Research Division (joined as assistant member, rising to head of genetics programme), Fred Hutchinson Cancer Research Center, Seattle, Washington  
**1991–93:** Staff scientist, Lawrence Berkeley National Laboratory, Berkeley, California  
**1990–91:** Postdoc University of California, Berkeley  
**1987–90:** Postdoc, Harvard University, Cambridge, Massachusetts  
**1987:** PhD Oregon Health and Science University, Portland